Mobility support for IP - based wireless networks

Mihailovic, Andrej

Centre for Telecommunications Research
King's College London
University of London
Strand, London WC2R 2LS
England

Mobility Support for IP-based Wireless Networks

By

Andrej Mihailovic B.Eng, M.Sc.

Thesis submitted to the University of London for the degree of

Doctor of Philosophy

Acknowledgments

I thank Professor Hamid Aghvami for his supervision and the trust he granted me during my research. He has given me the opportunity to develop professionally and gain confidence and optimism about my work.

My deepest gratitude is extended to all colleagues in the Centre for Telecommunications Research for the wonderful atmosphere during my research years. I also thank John Pearson for his help during the preparation of the text.

I thank my colleagues in British Telecom labs in Ipswich who helped me commence my work and supported me in realisation of my initial ideas. This especially goes to Dr. Mohammed Shabeer, Philip Eardley and Dave Wisely.

My appreciation goes to all colleagues from the European Commission projects, BRAIN and MIND, especially Tapio Suihko (NOKIA) and Nikolaos Georganopoulos (KCL) who worked on topics relevant to the thesis.

I express my sincere appreciation to my father Jovan, my mother Milica, my brother Luka and all of my family and friends for their support and encouragement. This work is dedicated to all of them. For the Glory of God.

Abstract

The main topic of this thesis is research into development of network layer mobility protocols for Internet networks. The main objective of this industrially-aware research was development of network layer mobility protocols. The results of the research are mainly applicable to specific types of Internet Protocol-based networks with wireless local area access technologies. These network environments were used as the basis in performance evaluation and consequent conclusions.

The thesis starts with the fundamentals of the Internet and its default support for mobility and justifies further research. A particular conceptual approach is used to explain the problem of mobility by introducing an abstract mobility problem statement and a new classification of mobility protocols. This continues with an investigation of Internet multicast as an alternative routing solution and includes an extensive analysis of all aspects of the possible adaptation of multicast for solving mobility. The focus is placed on the implementation flexibility of solutions. This converges in the proposed Multicast for Mobility Protocol, which is then tested using simulations and compared to other relevant protocols considering variety of parameters in small-scale networks.

The thesis then reflects on the conducted research and proposes a novel model for understanding, evaluating and enhancing mobility protocols in the Internet for both versions of the suite. This is represented in the Evaluation Framework, which is then applied practically in investigation of general mobility solutions. This results in new extensions to mobility design in the form of allowing creation of sub-protocols, called Protocol Design Issues Solutions, which can be deployed separately or integrated with existing mobility protocols, again achieving greater deployment flexibility. The actual sub-protocol proposed and covered in detail is a handover facilitating sub-protocol. This is concluded with the general outlook of possible applications of the approach to development of mobility protocols and examples of its practical realisation.

Table of Contents

Table of Figures

Table of Graphs

Chapter 4

List of Abbreviations

AAA:	Authentication, Authorisation and Accounting
AN:	Access Network
ANP:	Anchor Point
AP:	Access Point
AR:	Access Router
ARP:	Address Resolution Protocol
AS:	Autonomous System
BCMP:	BRAIN Compromise Mobility Protocol
BOOTP:	Bootstrap Protocol
BRAIN:	Broadband Radio Access for IP-based Networks
BS:	Base Station
CBT:	Core Based Trees
CCoA:	Collocated Care-of-address
CH:	Corresponding Host
CoA:	Care-of-address
DHCP:	Dynamic Host Configuration Protocol
DNS:	Domain Name System
DVMPR:	Distance Vector Multicast Routing Protocol
FA:	Foreign Agent
GGSN:	Gateway GPRS Support Node
GPRS:	General Packet Radio Service
GSM:	Global system for mobility
HA:	Home Agent
ICMP:	Internet Control Message Protocol
IGMP:	Internet Group Management Protocol

IP:	Internet Protocol
IPv4:	Internet Protocol version 4
IPv6:	Internet Protocol version 6
LD:	Location Directory
LERS:	Localised Enhanced-Routing Schemes
MANET:	Mobile Ad-hoc Networks
MER-TORA:	Mobile Enhanced Routing – Temporally Ordered Routing Algorithm
MH:	Mobile Host
MIND:	Mobile IP-based Network Development
M.IP:	Mobile IP
MMP:	Multicast for Mobility Protocol
MMPv4:	Multicast for Mobility Protocol for Internet Protocol version 4
MMPv6:	Multicast for Mobility Protocol for Internet Protocol version 6
MSC:	Mobile Switching Centre
MSM-IP:	Mobility Support for Multicasting in IP
MOSPF:	Multicast Open Shortest Path First
NAI:	Network Access Identifier
OSPF:	Open Shortest Path First
OSI:	Open System Interconnection
PAA:	Proxy Agent Architectures
PDI:	Protocol Design Issue
PIM:	Protocol Independent Multicast
PIM- DM:	Protocol Independent Multicast - Dense Mode
PIM-SM:	Protocol Independent Multicast - Sparse Mode
PPP:	Point-to-Point Protocol
PSTN:	Public Switched Telephone Network
RIP:	Routing Information Protocol
RARP:	Reverse Address Resolution Protocol

SGSN: Serving GPRS Support Node

SIP: Session Initiation Protocol

TCP: Transmission Control Protocol

UMTS: Universal Mobile Telecommunications System

CHAPTER ONE

The preliminary

This thesis presents results of research into mobility support for Internet networks by proposing a new mobility protocol called Multicast for Mobility Protocol, a novel model for developing generic mobility protocols and the resulting separation of functionalities and their mechanisms called Protocol Design Issue Solutions. A particular Handover Management Protocol Design Issue Solution is proposed and presented in detail. In addition, practical applications of this new generic mobility design are shown in the design of more recent mobility protocols, which include the Handover Management Solution presented here. The thesis is written using an original perspective on the development and analysis of Internet Protocol mechanisms for supporting mobility of Internet hosts where the emphasis is placed on research into efficient yet easily deployable mobility protocols for types of deployment scenarios considered in the thesis.

1.1 Motivation for the research

Internet technology has experienced an enormous expansion in recent times. The history of the Internet shows that its operational principles, as a widespread information infrastructure for accommodating heterogeneous networks, provide a basis for this substantial future relevance and growth. The commercial success of technologies supported by the TCP/IP protocol suite has also become the undisputed argument for its further development and inclusion in many telecommunications systems. It seems that the essence of the Internet network layer as the generic and "bonding" mechanism for underlying access mediums and higher layers, has brought this success. While this expansion is mostly related to the historic development of Internet networks for fixed communications, support for efficient mobility is undoubtedly one of the most important future milestones in the development of Internet technology. Beside mobility, Quality of Service support and secure Internet communications are other important topics, which again are dealt with in a specific manner in mobility situations.

An important progress in the popularity of Internet Protocol is noted in the ever-increasing consideration of the technology for current and future cellular telecommunication systems, which again are mostly concerned with providing services to mobile terminals.

Thus, mobility design for the Internet faces a dual task of developing features that will efficiently improve the default (and scarce) mobility support in the standard Internet and yet provide a solution that will be easily integrated in the applicable Internet deployment scenarios (see section 1.4.).

The research presented in this thesis is mostly concerned with expanding mobility support in the Internet with solutions that are, naturally, as efficient as possible but specific in the manner in which they are formed in the sense that

they enable easy realisation in Internet deployment scenarios considered in the research (see section 1.4). This again does not mean that the research in not concerned with pure Internet functionalities (as is the case with most of the document) but only that the striven solutions are aimed at having a practical side to their operation. This was mostly affected by the research background, which was heavily influenced by the industrial perspective on Internet development. The research specifically concentrates on the Internet routing, with a minimal dependence on underlying technologies as possible in order to preserve the generic nature of final solutions.

1.2 Introduction to TCP/IP protocol suite

TCP/IP suite is the collection of protocols facilitating the Internet. The Internet is the global networking infrastructure, which provides the communication medium for smaller local networks and hosts, also running the TCP/IP suite. TCP/IP is fundamentally a collection of the network layer, facilitated by IP [87], and the transport layers, i.e. Transmission Control Protocol - TCP [10] and User Datagram Protocol - UDP [11]. The two protocols are used to provide the platform for the application of other protocols: control or user based.

Internet Protocol (IP) is the network layer routing protocol of the Internet and provides the delivery mechanism for all packets belonging to different transport layers and applications residing on Internet hosts and routers (see Figure 1.1). IP "glues" the Internet together. IP provides a connectionless, "best effort" service and typically assumes that data security; order and delivery are controlled by higher layers (eg. TCP provides a connection-oriented transport). Internet bases its operation on the

distribution and assigning of IP addresses. An IP addresses[1], by definition, identifies an interface connected to the Internet.

An IP node (host or a router) usually has one IP address associated with each interface. Interface addressing is a more natural term when applied to routers which have multiple interfaces connected to different "sides" of the Internet thus requiring a different address for each interface. Although it is perfectly normal for an IP host to have more than one interfaces, and thus more IP addresses, those addresses are usually termed as the host's IP address[2]. There are several classes of Internet addresses as shown in Figure 1.2. The most importation distinction between addresses is the split between the unicast (Class A, B and C) and the mutlicast addresses (Class D) since they define a location dependent identifier and a group-based location independent identifier respectively (unicast addresses range from 0.0.0.0 to 223.255.255.255 and multicast ones range from 224.0.0.0 to 239.255.255.255).

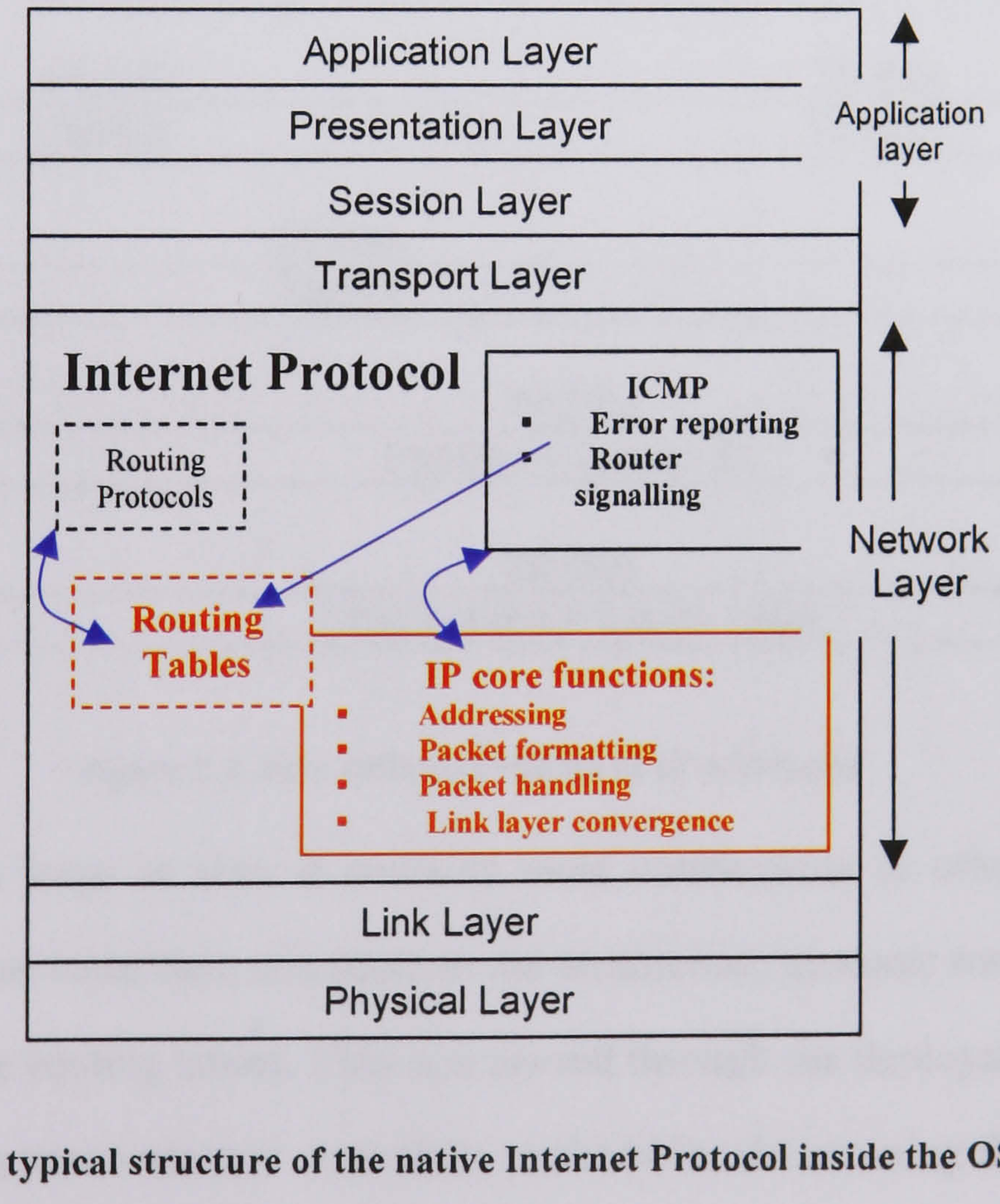


Figure 1.1. A typical structure of the native Internet Protocol inside the OSI layers.

---

[1] Internet Network Information Centre (InterNIC) assigns addresses to users (NetIDs).
[2] Again, it is possible to have a single-interfaced host with more than one IP address.

IP routing is one of the essential operations of the IP network layer and determines paths of IP packets from sources to destinations. IP routers normally perform packet routing[3] by deriving forwarding decisions from the internal routing tables. Routing tables are created in different ways. For small networks and attached local hosts, routing tables can be created in a *static* manner by manual configuration and can be assisted through the Internet Control Message Protocol [11]. The operation of using the routing tables is essentially a table processing procedure for determining the next hop of the received packet (usually a next hop router or a default router if the destination is not on the same link or subnet) by processing its destination address (see Figure 1.3).

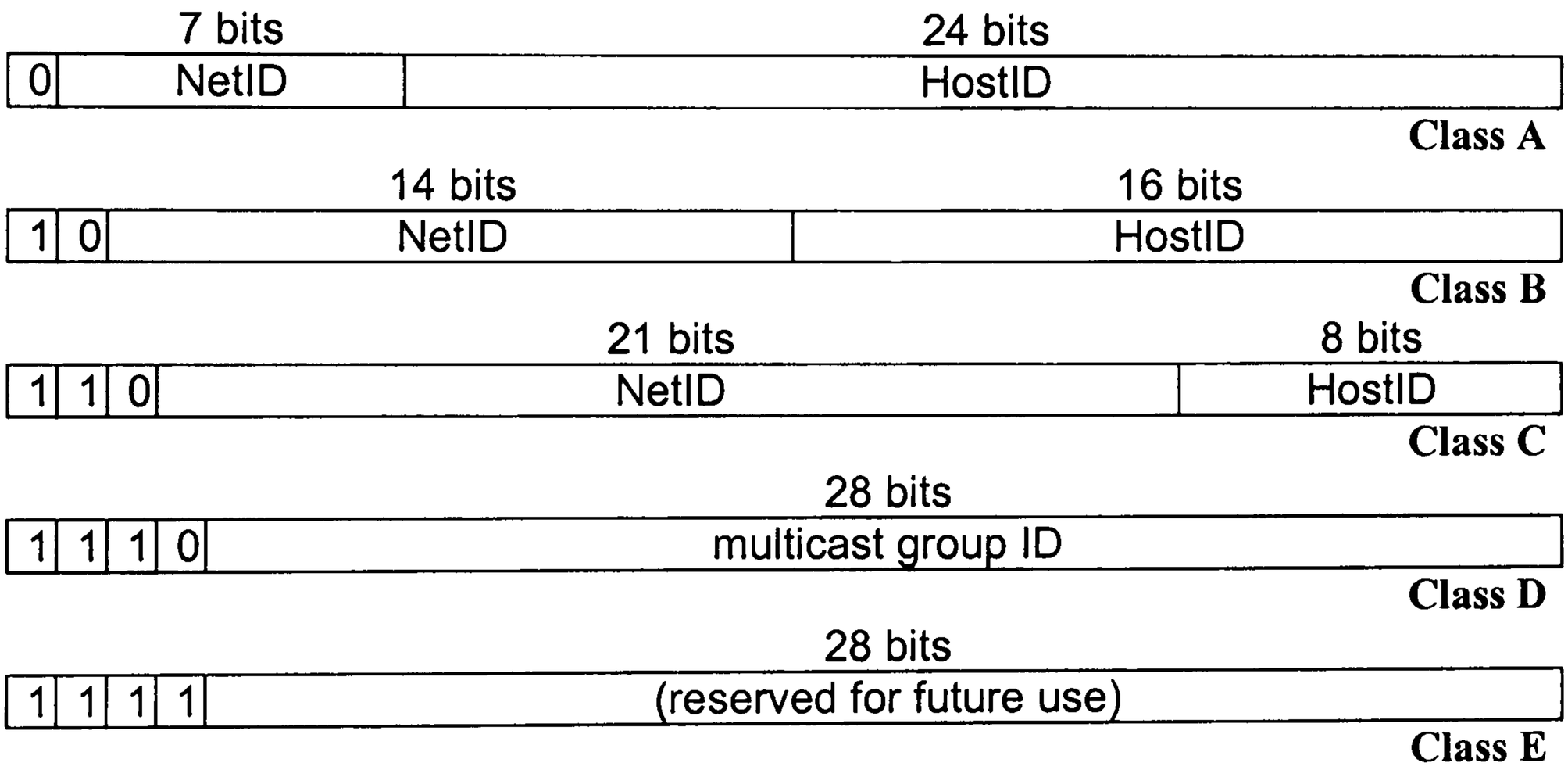


Figure 1.2. Five different classes of IP addresses

If the network is large in size, it contains more connections to other parts of the Internet and can use more than one route to the destination, *dynamic* routing is needed for configuring the routing tables. This is achieved through the deployment of routing protocols. Routing protocols rely on certain methods for determining the sequence of

[3] An IP host can also be configured as a router although restrictions may apply. The pure IP host should never forward any packets from one of its interfaces to another in the way routers do it.

routers and links that a packet traverses from source to destination. The method of determining a packet's path is the core operation of a routing protocol and is calculated by the routing algorithm (the path may be the same for all applied routing algorithms especially if the shortest path is the key criteria). There are two main approaches in determining the path of packets in routing algorithms: link-state and distance-vector routing algorithms expressed in Dijkasta and Bellman-Ford routing algorithms respectively. These two routing algorithms present an idealised platform for path calculations for the distance-vector based Routing Information Protocol - RIP [20] and the link-state based Open Shortest Path First – OSPF protocol [21]. Again, there are additional classifications of routing protocols in the Internet based on the scope of their deployment: inter autonomous system (AS) and intra autonomous system routing protocols. RIP and OSPF present intra-AS routing protocols deployable in collections of administratively scoped Internet networks. Inter-AS routing protocols are intended for routing between ingress routers of ASs and are more policy based than the cost-driven intra-AS routing (Border Gateway Protocol [45] being the base inter-AS protocol).

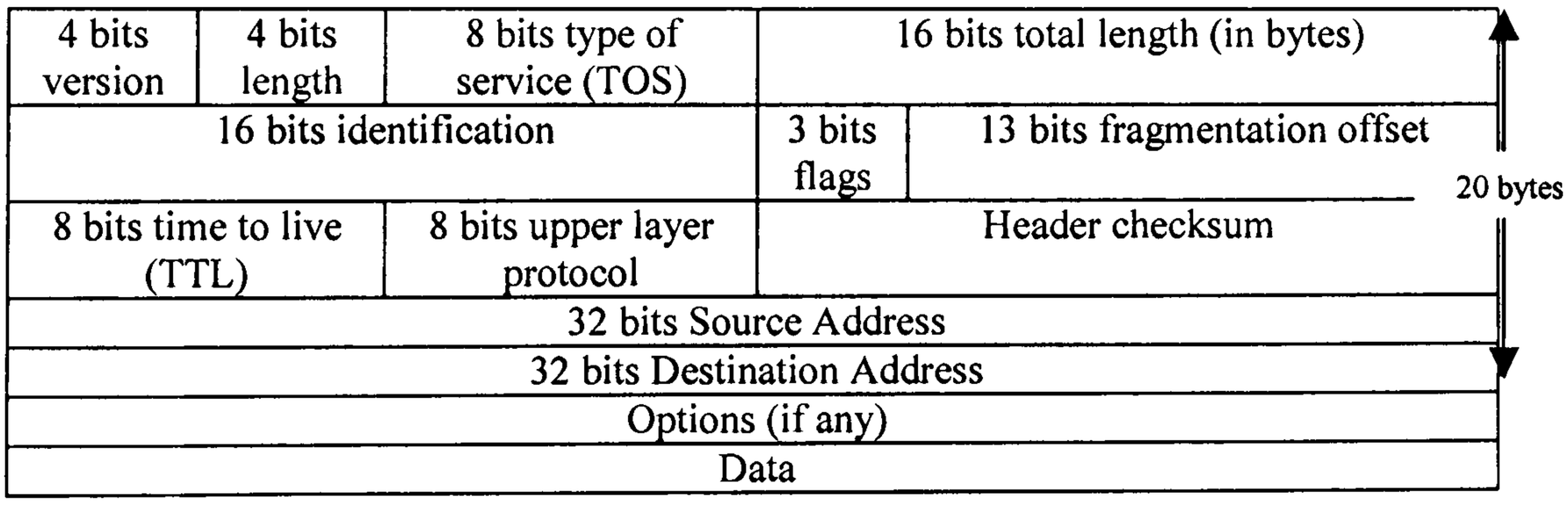


Figure 1.3. IPv4 header

Routing decisions in Internet routing protocols involve processing of the packets' destination address and determining the next hop in the path towards the network where the destination is located. This network is usually determined by the network part of the destination IP address, i.e. by the prefix of the IP address. IP networks can

be further divided using subnets and subnet prefixing. **Subnet** is a sub-network inside a prefix-determined IP network and provides means for dividing various collections of hosts (eg. connected to an Ethernet link) into routable portions of the original network. Subnets are usually transparent to the routing protocols outside the network, which is represented by the default network prefix of the IP address. This enables conservation of routing memory in the Internet since all routing entries can be aggregated into the original default network prefix and the remaining subnet routing can be performed transparently inside the actual network. Subnet addresses are created by reserving a portion of the hostID of an IP address for a subnet prefix, i.e. transforming a portion of the hostID into a subnetID. The remaining part of the hostID can then be used for allocation to individual hosts inside a particular subnet. Selection of the bits, which correspond to the subnet prefix is a decision made by the network administrator. For example: In a Class B IP address, 16 bits belonging to the hostID can be split up so that the initial 8 bits correspond to the subnet prefix (subnetID). In such a scenario, a single, class B network, can be further divided into 254 subnets with up to 254 hosts inside a single subnet.

1.3 Introduction to IP Multicast

Multicast is essentially a point-to-multipoint[4] communication between multicast group members where a group member can send and receive packets to and from all other members (in some cases a non-member can also send packets to a multicast group). An IP multicast group is a collection of Internet hosts which share the same Class D IP multicast address. Membership of the group is dynamic; hosts can enter or leave the group at any instant of the lifetime of the multicast session. A multicast IP address

is a flat ("prefix-less"), location-independent address, which ranges from 224.0.0.0 to 239.255.255.255. (in principle, any multicast group can use any available address from the pool of multicast addresses and advertise it as the identifier for its session) One block of multicast addresses, from 224.0.0.1 to 224.0.0.255 is assigned to various pre-arranged groups such as all routers on a subnet or all routing protocol-specific routers in a network. The pre-arranged groups are an efficient method of replacing bandwidth-consuming, trial-and-error Internet broadcasting. The remaining multicast address space is available for applications using multicast: either as permanently assigned addresses to multicast applications or for dynamic assignment. Additionally, a small range of multicast addresses, from 239.0.0.0 to 239.255.255.255, is assigned for locally scoped applications, which can be used in administrative domains and not necessarily in the global Internet.

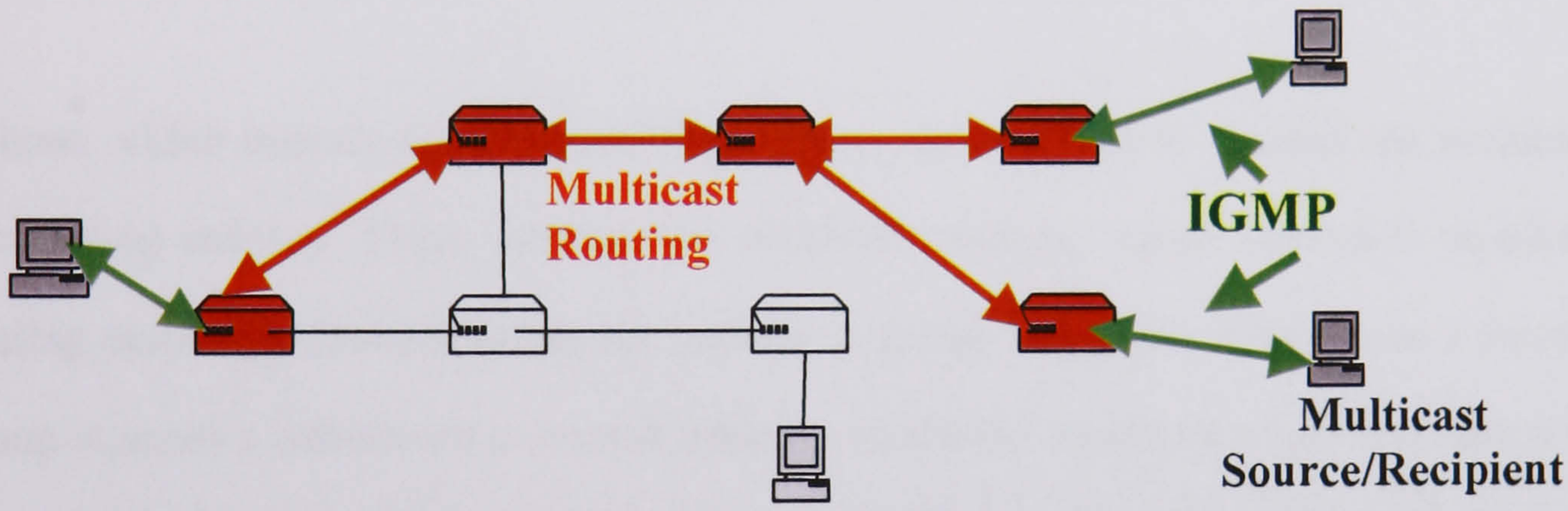


Figure 1.4. A typical setup in IP multicast

IP multicasting can be divided into two parts (see Figure 1.4):

- The first part is executed by potential group members including discovering the present group-to-session mappings (for example, through distributed session directories) and consequently contacting nearby multicast routers through the Internet Group Management Protocol (IGMP) [12].

---

[4] There are a mounting variety of Internet applications that can use multicast as an alternative to the bandwidth consuming unicast communication: video and audio conferencing, shared applications, information distribution (shares quotes…).

- The second part involves setting up of forwarding routes and delivery of packets to destinations. This is accomplished by multicast routing protocols: DVMRP [15], MOSPF [16], PIM-dense (DM) and PIM-sparse mode (SM) [17] and Core Based Trees (CBT) [18]. Multicast routing protocols establish multicast routing trees in Internet networks as collections of routing entries in involved routers. The result of this is that all group members are "connected" and packets get distributed to all members. Similar to the unicast routing protocol explained in the previous section, multicast routing protocols construct routing tables for forwarding received packet. Forwarding of packets in a multicast protocol is always performed away from the source; i.e. the packet should never come back towards its origin.

1.3.1 Internet Group Management Protocol - IGMP

A host, which intends to join a multicast group, needs a way to contact the multicast facilitating entities. These entities are multicast routers, which construct multicast routing trees and forward multicast packets to group members. When a host joins a group it sends a membership request message to a local multicast router and sets it IP process and network interface card to receive packets addressed to that group. This process is handled by IGMP [12] by providing mechanisms for Internet hosts to contact and inform neighbouring multicast routers about the relevant multicast group(s). Multicast routers use this knowledge to join relevant multicast routing trees and participate in packet forwarding on behalf of the local host[5].

The first version of IGMP [12] defines protocol procedures for achieving the communication between multicast group members and local multicast routers. Essentially, the process consists of exchanging and processing IGMP controlled messages: Reports and Queries transmitted by hosts and multicast routers

respectively. The second version of IGMP [13] enhances the first version with some new features such as the new *Leave* message used by group members, when they want to cancel group membership. This message provides for explicit release of group membership and reduces the "leave latency" present in the first version of IGMP. This "leave latency" in the first version of IGMP was caused by the requirement for a timeout for determining when a host has terminated its group membership. IGMP version 3 [14] adds some new features such as filtering of multicast traffic by including/excluding traffic (i.e. packets) from particular group members.

1.3.2 Multicast Routing Protocols

Besides facilitating the first step in multicast communication, IGMP provides for the final delivery of packets from the local router to the attached multicast group member[6]. The distribution of packets across Internet networks to the "leaf" of the multicast routing tree is performed by multicast routing protocols. Multicast routing protocols are responsible for the construction of routing entries in routers belonging to multicast trees and forwarding of packets to group members. Various algorithms (techniques) can be used for the construction of multicast trees and delivery of packets. The most explanatory classification of multicast protocols, which defines the essential differences resulting from the applied routing algorithms and protocol mechanisms, separates the protocols into two types:

a) ***Dense mode*** multicast routing protocols: *Dense mode* protocols are designed to work well in densely populated environments where there is a multitude of group members. This is mainly due to the fact that *dense mode* protocols deploy routing

---

[5] A standard assumption is that multicast group members are at the "leaf" of the multicast tree, meaning they have no further downstream routes/routers.

[6] Local/neighbouring multicast router is assumed to have either physical or logical layer 2 connection with the multicast host running the IGMP.

algorithms, which are not scalable in situations where group members are sparsely distributed across the network at various locations. There are three main examples of *dense mode* multicast routing protocols:

- **Distance Vector Multicast Routing Protocol (DVMRP)**: DVMPR constructs a multicast tree for every source of multicast packets. The algorithm for constructing multicast trees and router entries is a simple "broadcast and prune" (see Figure 1.5) technique. Sources transmit multicast packets, which are forwarded by multicast routers towards "leaf" networks. Forwarding is based on source checking: all multicast routers receiving packets construct a routing table to validate that the packet received from a certain source was actually received from the incoming interface for that source (RIP is used for determining the incoming interface). If the packet's incoming interface is validated, the router stores the (***multicast group, source***) entry in its routing table and broadcasts the packet on the outgoing interfaces for that source. Multicast trees "shape up" when "leaf" networks, which do not have group members (determined by IGMP), send a prune message in the opposite direction and eliminate their particular routing branch from the tree. DVMPR further defines mechanisms for "grafting" previously pruned tree branches and maintaining the routing entries.
- **Multicast Open Shortest Path First (MOSPF)**: MOSPF is built as a multicast extension of the OSPF routing protocol. The essence of the operation is that for every source of multicast packets (***multicast group, source***) an entry is created in the router by using the link-state provided in MOSPF routers. Unlike DVMRP "broadcast and prune" of multicast packets is not used to construct a multicast tree but to provide predetermined information about group members to which multicast packets need to be distributed. This information is **flooded** in the form of a link-state update by local multicast routers of group members attached to the whole network.

Thus MOSPF is a source-based multicast routing protocol which uses "explicit joining" of group members for constructing multicast trees. Additionally, MOSPF builds multicast trees "on-demand", only when a router receives a first packet from a multicast source. MOSPF defines some additional mechanisms for providing multicast in different network scenarios but is mostly restricted to OSPF-using networks.

- **Protocol Independent Multicast[7] – Dense Mode (PIM-DM)**: PIM-DM deploys a similar "broadcast and prune" mechanism as DVMPR but is independent of the need for any underlying unicast routing protocol to perform the incoming interface check for sources of multicast packets.

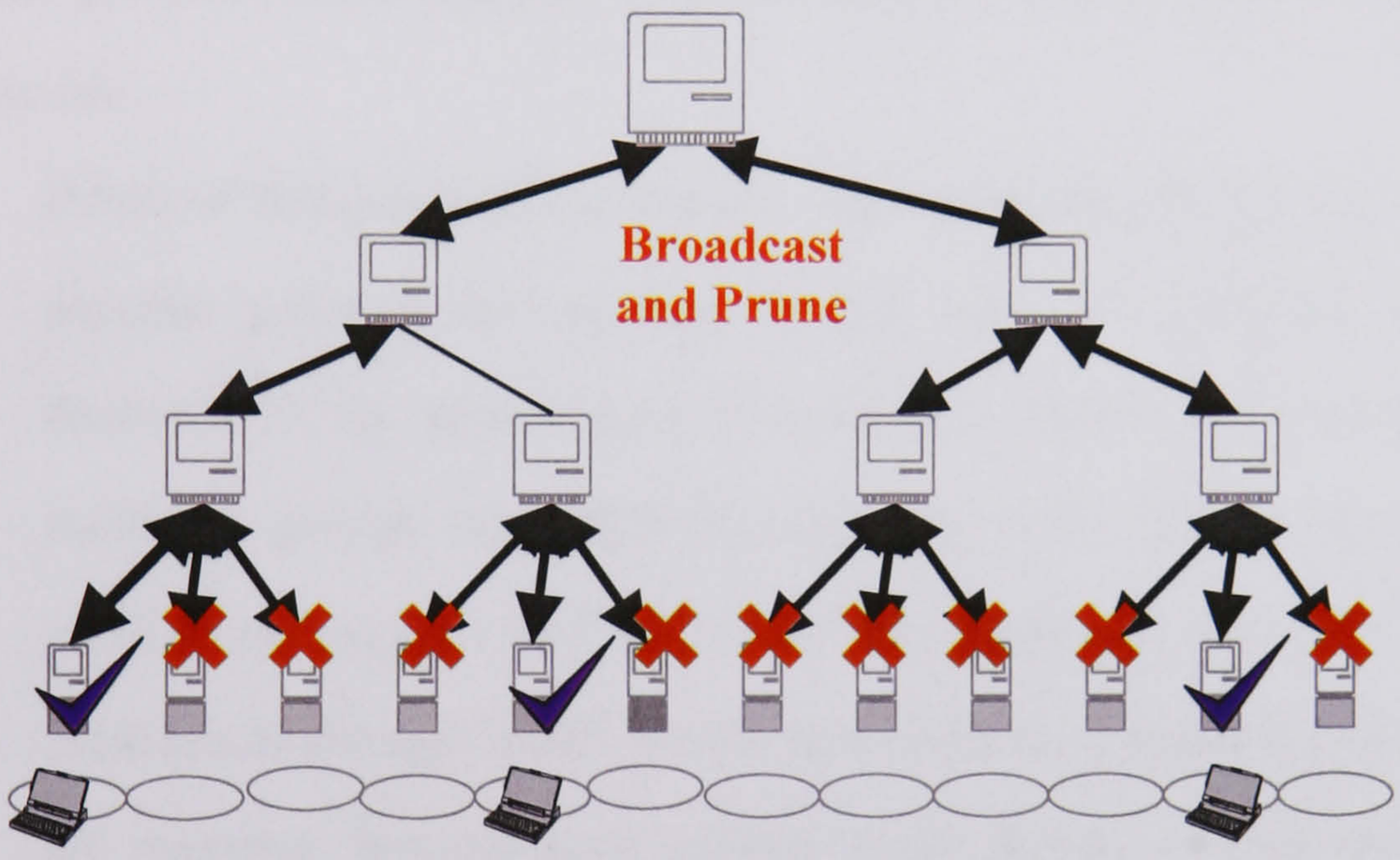


Figure 1.5. A *Dense Mode* multicast scenario

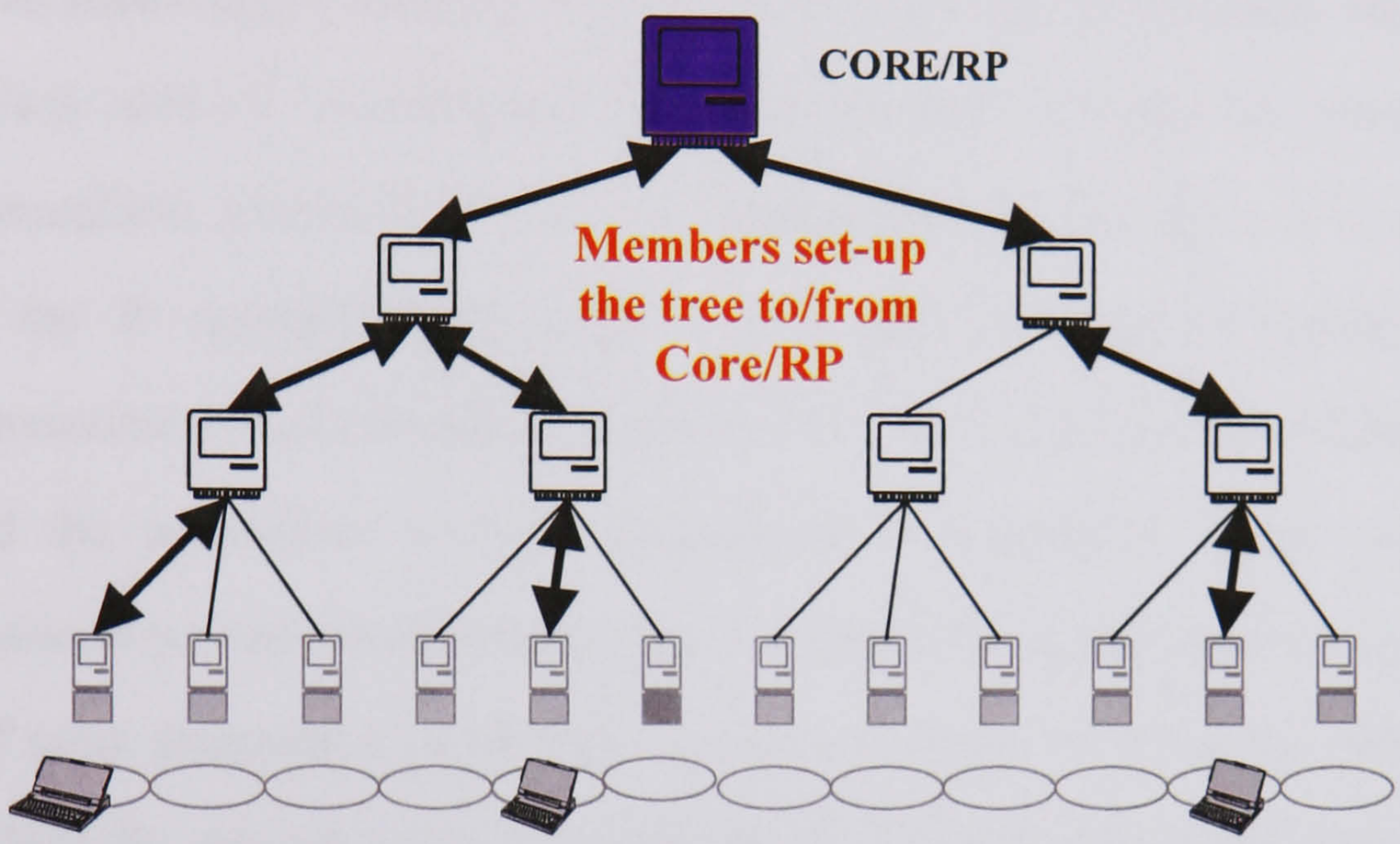
[7] PIM (both Dense and Sparse mode) is designed to be used in any Internet environment hence it is applicable to both intra-AS and inter-AS systems. While DVMRP and MOSPF can appear to be restricted to mostly intra-AS environments due to the limitation imposed by the supporting unicast protocol (both RIP and OSPF are mainly intra-AS protocols), they can also be applicable for inter-AS networks (current research concentrates on this). A particular example of this in the Internet Multicast Backbone (Mbone): the global interconnection of multicast capable routers for delivery of IP multicast. Mbone however, sees tunnels instead of interfaces as the outgoing "pointer" for multicast packets due to the lack of multicast routers in the Internet.

b) ***Sparse mode***[8] multicast routing protocols (see Figure 1.6): *Sparse mode* protocols are primarily intended for scenarios where multicast group members are widely dispersed across locations spanning several networks. In these situations, efficiency of *dense mode* protocols due to the simple flooding-based protocol mechanisms is not the optimum solution. In fact, a protocol should minimise the impact of control procedures in such way that bandwidth is preserved and that there are no unnecessary distributions of packets across networks. The key operation of *sparse mode* multicast routing protocols is the presence of a central, focal router for each group to which packets are sent from multicast sources and which recipient members "explicitly join". The packets are then distributed along the routes formed by the protocol. There are two main examples of *sparse mode* multicast routing protocols:

- **Protocol Independent Multicast – Sparse Mode (PIM-SM)**: PIM-SM is a separate protocol for multicast routing from the PIM-DM and is never deployed in the same region of multicast routers. The central routers for multicast groups are called Rendezvous Points (RPs). Group members perform an "explicit join" to the RP for a particular group by contacting a local router through IGMP, which then performs a discovery of the group-to-RP mapping. Sources send packets to the RP by making the local router encapsulate multicast packets into a special message addressed to the RP. The RP can then decide to join the particular source to the group and create a forwarding state on the path between the source's local router and a RP. PIM-SM additionally allows some receivers to switch to a source-based tree for a particular source. This is performed by sending an explicit message to the source and forming a source-based tree. In this scenario, some traffic is diverted from the RP.

[8] Sparse does not imply that the multicast group has fewer members; it only indicates that they are more distributed.

- **Core Based Trees (CBT)**: CBT deploys a simpler mechanism to PIM-SM and does not include the possibility of switching to source-based trees for a subset of the group. The central routers of multicast groups are called Cores. The operation of CBT is symmetric, both sources and receivers in a multicast group join the Core and create a multicast routing tree for the group. The traffic is then distributed unidirectionally from sources and receivers.



Multicast group members

Figure 1.6. A *Sparse Mode* multicast scenario

1.4 Scope of the Research

IP network layer is applied in various types of telecommunication networks facilitating access of IP terminals or transit of IP traffic. Factors such as different wireless access technologies, administrative issues, network sizes, terminal populations, network performance parameters and accompanying network infrastructure for both higher and lower layers can be used to further characterise different IP networks. For this purpose it is beneficial to define network scenarios that are used in the here-presented research and hence define the direct relevance of the

results and the scope of research. These network scenarios are considered to identify IP networks to which the research is applicable. Selection of the used network scenarios and IP networks they represent is explained in the following:

a) Initial IP mobility network scenarios: Starting from the first widely accepted mobility protocol Mobile IP [28] by Internet Engineering Task Force (IETF), the general approach with many other mobility protocols [29-41][83-86] is to abstract three main logical entities involved in mobility of IP terminals: home network, visited network (sometimes referred to as home and foreign sites as done in Hierarchical Mobile IP [33]) and the global Internet. The focus is on development of the IP network layer features. One such developed network scenario is represented in the network topology used in section 3.4 for development of MMP and the parameters used for simulations in Chapter 4. Some other mobility protocols provide similar descriptions for network scenarios and topologies using the same abstraction of the three network entities. For example, Cellular IP [35] defines the network scenario applied where in Cellular IP global Internet is named Internet backbone with Mobile IP and foreign and home networks are Cellular IP networks with Gateways used as ingress/egress points of the networks. Furthermore, Cellular IP generalises functionality of the wireless link technologies by defining bases stations as IP routers, which periodically emit wireless beacons. Fast and Scalable Handoffs [34], Deadalus [38] and MSM-IP proposal [39] test their protocols in IP networks with Wireless Local Area Network (WLAN) technology. HAWAII [36] details properties of the network scenarios used in the investigations with similar characteristics. Regarding the network performance parameters used in simulations throughout the thesis, these again are following this approach in referenced mobility protocols [38][39][40][47][82-86] discussed in Chapter 4.

b) Broadband Radio Access for Internet Protocol Networks (BRAIN) approach [47]: The BRAIN project was a European Commission sponsored project with a

consortium of telecommunication companies and universities developing IP technology (more details are given in Appendix 2 also explaining the follow-up project MIND). One of the main goals of the project was design of an IP-based mobile wireless access system. The first step in realising this goal was definition of the IP wireless access network (or more appropriately, BRAIN wireless access network), which would then be used for extracting network scenarios for development of other functionality such as mobility protocols. Considering the impact on mobility and relation to the network scenarios mentioned under a) one of the architectural conclusions of the project was that the network scenarios considered in the mobility protocols mentioned above under a) can be used as a valid abstraction of BRAIN IP access network and therefore applied in the mobility research. There are several conclusions, which justify this approach. The first conclusion from the BRAIN project is that the network in consideration is fully IP-based including the Gateways and access routers (economic, engineering and scientific reason for this conclusion are included in the project deliverable [47] mainly emphasising the need for capitalising on the current success of the Internet and "pushing the IP router to the very edge of the terrestrial network, a single hop from the mobile user"). Some BRAIN design principles, which further explain the meaning of IP-based in the BRAIN context and the approach taken for development of IP mobility protocols are [47]:

1. *Obedience of the Internet end-to-end principle*, which was considered to encompass the following requirements: independence of uppers layers and independence on types of IP packets being transported.
2. Another design principle was the *layered design* approach using IP network layer as the focus and "assuming independence and interoperability with upper layers and generic interface towards the link layers such that new and old link layers can be exploited without redesigning the network infrastructure" (see Figure 1.1.).

3. *Modularity* assuming that the IP wireless access network can fit into the existing Internet infrastructure considering the default functionality such as routing and addressing and support for functionality such as mobility protocols (i.e. micro and macro mobility interworking, see next chapter).

The second conclusion for the BRAIN project, being the direct consequence of the above discussion, is that the investigation of mobility protocols in BRAIN should follow the same approach as applied in other mobility protocols mentioned above under a) considering both network scenarios [5][8][47] and performance evaluations (see Chapter 6 for simulation parameters and [47][61]) and most importantly using the research conclusions from the mobility protocol development and testing, which applied the approach explained above under a). Note: regarding the size of the BRAIN access network, this was arbitrary. However, performance evaluation uses network sizes and terminal population similar to other mobility protocols mentioned under a) (see Chapter 4, Chapter 6 and [47][61]). In addition, BRAIN project considered WLAN-type wireless access scenarios, which were applied in the test-bed evaluation of the project's results (in addition to HyperLAN 2 which was researched in the project).

This analysis is intended to explain the approach used throughout the thesis and identifies the scope of the research based on the considered network scenarios. The scope can be further characterised with two main targets areas in mind:

- Pure IP network layer mechanisms for solving mobility, i.e. features of IP mobility protocols
- **Considered network scenarios map into IP networks, which use local area wireless link technologies and sizes corresponding to the campus area Internet networks.** While the first target can be realised in various IP networks of arbitrary using the abstraction previously descried, this target is influenced by the particular scenarios used in performance analysis in the thesis.

This leads to the conclusion that the results of the research shown in the thesis cannot be entirely applicable to 2.5 and 3G cellular telecommunication networks, which are also deploying IP technologies. The particular parameters of cellular network were not considered in the design stages nor in the performance analysis. This is seen as one of the major items for future research as indicated in Chapter 7.

In the general attempt to make this research as broadly applicable as possible, there are several factors that could be considered to assist in deducing a rough estimate on the applicability of mobility solutions developed for any new IP network:

- IP network layer as the focus with minimal dependency on upper and lower layer protocols. This may not provide enough confidence in the applicability to all network environments but at the same time releases dependency on any specific network scenario apart from the general ones explained above. The level of abstraction offered by the above network scenarios could be used for abstracting different Internet networks considering the functionality of IP network layer for solving mobility.
- Some simulation conclusions could be expanded to any network scenario. This may be generally extracted for handover performances of mobility protocol (Chapter 4 and Chapter 6) while it may be more difficult for protocol overhead and scalability study which may require larger networks and population of terminals (analysis in Chapter 4 attempts to provide conclusions for such scenarios)
- Generic Mobility Design Models presented in Chapter 5 shows a model for applying specific design criteria in the process of creating mobility protocols. While this models is practically applied in the BRAIN project with the specific design principles but not different network scenarios (including sizes and terminal populations) this experience can be useful for investigating other IP networks.

1.5 Thesis outline

Chapter 1: This Chapter gives the motivation behind the research conducted to indicate the preliminary perspective on what the final solutions are aiming to achieve. The Chapter concludes with a general introduction to the basic features of Internet Protocol, which are used as a foundation for the subsequent parts of the document. The thesis starts with an overview of Internet Protocol version 4 in Chapter 1. Later on, the remaining Chapters introduce the awareness of the new Internet Protocol version 6 and analyse all mobility aspects bearing in mind the differences between the two versions.

Chapter 2: In Chapter 2, a detailed insight into mobility in the Internet is given. It elaborates on the particular approach to mobility used in this research and defines the exact problem of host mobility in the Internet. A particular analysis of mobility protocols is given and a conceptual representation of the problem is identified through an abstract mobility model. The Chapter concludes with a new classification of mobility protocols along with the justification for the classification criteria applied. The classification includes all mobility protocols relevant at the time of writing of the thesis. It disregards the chronological order since some protocol emerged significantly earlier that others. This is because the relevance is placed on distinguishing the mobility approaches and showing the evolution in the conceptual approach to the problem. This also applies to the protocol design shown in Chapter 3, which was one of the earliest mobility proposals.

Chapter 3: Full steps of the design of a mobility protocol, called Multicast for Mobility Protocol, are given in Chapter 3. The Chapter starts with an analysis of features of multicast in the Internet and their potential adaptation for solving mobility. Three possible models for integration of multicast for solving mobility are outlined and reasons given for selecting the particular model in the design of Multicast for

Mobility Protocol. The work then presents all features needed for completing the development of the mobility protocol along with a thorough description of all the operational procedures and the need for alteration of the existing Internet architecture required to support the operation of the protocol. The end of the Chapter contains a brief introduction to Internet Protocol version 6 and an adaptation of Multicast for Mobility Protocol for functioning in the new version of the Internet Protocol.

Chapter 4: Chapter 4 shows the simulation of Multicast for Mobility Protocol and gives a comparison with some relevant mobility protocols considering handover performance and protocol overhead as the key parameters. A detailed analysis of the simulation strategy is presented along with the description and validation of the simulation results shown and additional mathematical analysis.

Chapter 5: Chapter 5 reflects on the topics covered in the previous chapters and presents a new practical generic model for evaluation and development of mobility protocols called the Evaluation Framework. The model was developed with an industrially-aware research perspective for creating mobility protocols with greater deployment flexibility but still satisfying the standard efficiency criteria of Internet mobility protocols. This culminates in the split up of mobility mechanisms and their functioning solutions called Protocol Design Issues and Solutions respectively. The model proposes a broken-down approach to mobility design represented by the Protocol Design Issues identified. Example applications of the model are shown in the creation of BRAIN Candidate Mobility Protocol (developed in BRAIN/MIND projects) and the enhancement of Multicast for Mobility Protocol.

Chapter 6: In Chapter 6, results from the generic design model for mobility protocols are shown in the design of Handover Management Solution. The design follows the recommendation from the Chapter 5 and present relevant and detailed analysis of all topics in the handover design. Two types of handover are proposed, planned and unplanned, where planned handover is the complete process partially performed by the unplanned handover. The whole design is centred on achieving generic and easily

deployable solutions. To show the actual realisation of the proposed novel mobility design shown in Chapter 5 and Chapter 6, two examples are taken to show how the Handover Management Protocol Design Issue is incorporated with the rest of the mobility functions to form a fully functional mobility protocol. One example shows how an existing mobility protocol, in particular Multicast for Mobility Protocol, could be enhanced with the Handover Management "sub-protocol". Another example shows a protocol called BRAIN Compromise Mobility Protocol, which is designed entirely obeying the generic mobility design model shown in Chapter 5 and the inclusion of Handover Management in the protocol. The chapter includes simulations and performance analysis of the proposed Handover Management protocol.

Chapter 7: Chapter 7 gives a summary of the work presented in the thesis and an overview of what the research conducted managed to achieve along with some suggestion for future work.

Appendix 1: Appendix 1 shows the continuation of Hanover Management design by presenting the complete operation of the "sub-protocol" with the exact specification of the transfer and operation of control messages.

Appendix 2: In Appendix 2, details of the project BRAIN and MIND are included (MIND was the follow-up project of BRAIN). Author of the thesis participated in both projects and some of the work presented in this thesis was also included in the activities of the projects (see section 1.6).

Appendix 3: Appendix 3 shows a graphical representation of the protocol steps of the Multicast for Mobility Protocol to assist the description of the protocol in Chapter 3.

Appendix 4: Description of ns-2 is included in Appendix 4 along with some further simulation results from the BRAIN/MIND projects.

The results of the research can be divided into three distinct items, which are intended as the unique contribution of the thesis. They are listed below with their basic interrelations (more on their interrelations is given in relevant chapters):

a) **Multicast for Mobility Protocol (MMP)**: MMP is designed in Chapter 3 and its testing and analysis is in Chapter 4 [1][2]. MMP is further considered in Chapters 5 and 6 using the results of the Generic Mobility Design Models and application of the Handover Management protocol as a sub-protocol patch.

b) **Evaluation Framework for IP mobility protocols, Protocol Design Issues (PDI) split for IP mobility protocols as an element of the Evaluation Framework and its application for construction of IP mobility mechanisms constituting the Generic Mobility Design Model**: One of the conclusions from the development of MMP and other considered mobility protocols is that there is a general trade-off associated with their performances. One of these trade-offs is shown in Chapter 4 where MMP is improving handover latencies at the expense of increased protocol overhead. This quantitative trade-off can be further expanded with multitude of descriptive design criteria. This was the starting point for development of the Evaluation Framework [5], which reflects the need for an effective model for evaluating mobility protocols including all relevant parameters both qualitative and quantitative. Evaluation Framework applies modularity for analysis of mobility protocols and splits mobility functions into 9 PDIs, which are separately analysed considering their interrelations and impact in the design process. These conclusions are used as the foundation for proposing the Generic Mobility Design Model [8]. The model is intended as a tool for analysing and constructing mobility protocols with respect to specific design criteria, which may be chosen subjectively, for example, based on deployment requirements. This is demonstrated in the BRAIN project mobility platform consisting of 5 PDIs, which form the skeleton of any mobility solution conforming to them. One such protocol is BRAIN Candidate Mobility Protocol (BCMP was developed in BRAIN/MIND projects) and this is used to demonstrate the application of the Model. Chapter 5 explains how MMP can be adjusted to conform to the BRAIN platform and details a model for inclusion of Handover Management PDI with MMP and BMCP.

c) **Design of Handover Management protocol as a specific PDI**: Handover Management protocol comes as a direct consequence of the Generic Mobility Design Model and modular design of mobility protocol. Some modules, i.e. PDIs can be separated to such extent to be designed as separate sub-protocols as done for the Handover Management PDI designed in Chapter 6. It was intended to compliment any mobility protocol that conforms to the PDI split developed in Chapter 5 [8]. Handover Management protocol can be used as an enhancement to existing mobility protocols adjusted to the PDI split as shown for MMP and BCMP in Chapter 6.

The thesis follows the chronological ordering of the research and shows the evolution in IP mobility design that the thesis aims to reflect. This is especially applicable to the Generic Mobility Design Model where the goal of is not a perfect mobility solution but a practical compromise for a given deployment scenario. This is seen an important research conclusion. Few additional issues have influenced the structure of the thesis:

- MMP is designed following the general approach for development of IP mobility and offers useful results especially when compared to other IP mobility protocols
- Generic Mobility Design Model is deployment-centric and aims at providing a tool for development of mobility protocols with specific design principles (as used in the BRAIN project). Design of MMP was void of that specific deployment parameters and follows the general applicability also used in the outside research.
- Experience and conclusions derived from the development of MMP are the foundation for the Generic Mobility Design Model, without this its development would not have been possible hence it also serves the purpose of justifying the approach taken after the MMP's creation

From the above is can be observed that MMP is not being promoted as the undisputed mobility solution. This is explained in Chapter 4 where MMP is compared to other mobility protocols and where its advantages and disadvantages are revealed.

Chapter 5 introduces BCMP to show the realisation of the Generic Mobility Design Model in the BRAIN project. BCMP applies the Handover Management PDI designed in Chapter 6 thus introduction of BCMP is considered useful for explanatory purposes (BCMP is not attributed to the author).

Handover Management protocol is designed separately in Chapter 6 after Chapter 5 concludes that some mobility features can be designed as separated modules and added to exiting protocols. This is analysed for MMP and BCMP. Performance analysis is done when the Handover Management protocol is integrated with BCMP but the results of simulation are focused on the performance of the Handover Management protocol and they are related to both MMP and BCMP.

Two distinct sets of simulations are shown: one for MMP and one for the Handover Management protocol using OPNET modeller (Chapter 4) and ns-2 (Chapter 6 and Appendix 4 including the description of ns-2) respectively where both simulation tools are considered appropriate for network layer testing. MMP was tested solely by the author, whereas the Handover Management was tested in the BRAIN/MIND projects where ns-2 was considered more appropriate due to its public availability.

1.6 Research History of the presented BRAIN/MIND Project Results

Chapter 5 and Chapter 6 (BRAIN and MIND projects are also covered in Appendix 2 and 4) contain results and reference to the work conducted in BRAIN and MIND projects. The author's contributions are presented in the following paragraphs along with further description of the projects' issues and an explanation of joint project activities that are included in the thesis and are relevant to the presented results:

- As mentioned in Chapter 5 and 6, the work presented is mostly dealing with scope and topics of the BRAIN project and its refinements and testing in the MIND

follow-up project. This is not related to the additional scope of the MIND project related to additional network scenarios and resulting technical issues. Work referenced in Chapters 5 and 6 is related to work conducted in projects' Work Package 2 which dealt with networking issues and is contained in referenced final public deliverables 2.2: "BRAIN architecture specifications and models, BRAIN functionality and protocol specification" from the BRAIN project [47] and "MIND protocols and mechanisms specification, simulation and validation" from the MIND project [89].

- The author was a full time member of the BRAIN and MIND projects from King's College London working in Work Package 2 in both projects. King's College London was the overall leader of Work Package 2.
- The basis for Chapter 5 and in particular section 5.3 is initially presented as "experience of BRAIN and MIND project in development of IP mobility solutions" and was put forward to the Internet Research Task Force (IRTF) of Internet Engineering Task Force (IETF) as "draft-mihailovic-brain-mind-00" showing principles of the model that can be applied in development of IP mobility solutions. This is contained in personal reference [8] co-authored[9] with Mark West (Siemens/Roke Manor Research), Robert Hancock (Siemens/Roke Manor Research), Philip Eardley (British Telecom) and Tapio Suihko (Nokia). The document was created at the end of the MIND project recollecting and summarising the whole research and results on targeted IP mobility solutions from the projects (all authors of [8] were the major contributors to the overall mobility management work in the BRAIN project).
- Principles of the Evaluation Framework as contained in Chapter 5 are initially presented in personal reference [5] co-authored with Philip Eardley (British Telecom) and Tapio Suihko (Nokia). Concepts of the Evaluation Framework (in

[9] The employers of the co-authors mentioned are taken from their employment during the projects. The full names of the project consortium members, as included in the project, are contained in section A2.1.

particular PDIs and evaluation criteria) are also contained in personal references [3][90][91] from the BRAIN project where the co-authors were dealing with QoS interaction issues with mobility protocols. Another application of the basic principles of the Evaluation Framework is presented in personal reference [92]. Work presented in Chapter 6 follows the principles of PDI split in mobility protocols and its design and is based on results of the joint effort on mobility management in the BRAIN project mostly produced by all authors of [5] and [8].

- BCMP is not the author's contribution as mentioned in the thesis. BCMP is proposed inside the BRAIN project and included as one of its results, which are also discussed in Chapter 5 and Chapter 6. By being one of the results of the BRIAN project, BCMP was further developed in MIND project considering its broadened scope and further refinement and testing of its initial performances. Some of the work on general mobility management adaptation for multi-homing including BCMP (such as its Path Updates and Handover Management features...) is contained in personal references [94][95] created in the MIND project. In addition, BCMP is currently used for various further researches outside the BRAIN and MIND projects. One such example is consideration of BCMP and other mobility protocols with QoS issues contained in personal references [96][97].
- The simulation results and collective work on performance evaluations presented in Chapter 6 is not attributed to the author but is a result of the joint work performed in the projects and is taken from the work presented in public deliverables of the projects [47][89].
- Regarding the joint results of the project included in the thesis and individual contributions of other project members, the author highlights contributions (see Acknowledgements) by Nikolaos Georganopoulos (King's College London) who was the principal contributor of the simulations presented in the Chapter 6 and Tapio Suihko (Nokia) who was the principle contributor of Handover Management specifications presented in Appendix 1.

CHAPTER TWO

Introduction to IP Mobility

Chapter Overview

This chapter introduces the main research issues that are used as foundation for work presented in the rest of the thesis. Emphasis is placed on host mobility and the chapter includes a brief description and differentiation of user and host mobility concepts, which are used as the starting point for describing reasons for providing mobility in the Internet and constraints imposed by the default IP functionality. This is summarised in the issues that constitute the mobility problem in the Internet continued with a description of Mobile IP being the basic and reference IP mobility solution. The chapter concludes the study of IP mobility essentials by proposing a conceptual representation of the problem summarised in the abstract mobility model which is then used for understanding different concepts applied in more complex IP mobility protocols and provides a basis for the proposed classification of IP mobility protocols and abstraction of the problem applied in the remaining chapters.

2.1 Mobility Concepts

There are ever-expanding varieties of host scenarios in the Internet. This statement applies to various aspects of Internet connectivity: available applications, diversity of terminal equipment, heterogeneous nature of Internet networks, flexibility in host behaviours…This large range of functionalities, supported in the Internet, results in different mobility scenarios, which can also be separated conceptually. The starting point in the attempt to distinguish between different mobility concepts are the entities, which require mobility support in the Internet. These entities can be categorised into two types:

a) ***User;*** *User mobility* is a general concept of providing a set of functions, which allow a user to obtain access to the services provided by a network. The user is assumed to be able to connect to any terminal and achieve a form of network access. The functions facilitating a user's mobility usually include exchanges of signalling messages and achieving virtual connectivity from user to any required entity in the Internet[1]. In order to allow a particular user to obtain access to desired services there have to be mechanisms for: allowing access, providing local identity, configuring the network to supply the user-specific services and, most importantly, keeping the user's current identity in locations available to the rest of the Internet. While this whole set of functions can be recognised as *user mobility,* a particularly important aspect is the mechanisms required for the distribution of user's identity for the purpose of allowing session establishments. This is recognised as ***personal mobility***. It should be possible to establish a session regardless of a user's current location, the location being an IP host that a user may be using at any instant. Obviously, there is a split between the identifiers for IP hosts and users; these are IP addresses and specific identifiers[2] respectively. Distribution of a user's identity

[1] Usually this refers to users' home domains (company, house…).
[2] Examples of these identifiers could include: NAI (Network Access Identifier), DNS name, Caller ID (for SIP), Email address…

is mostly concerned with performing some identifier-mapping mechanisms either to a particular IP address or a local server or proxy (examples include Domain Name System- DNS [23] and Session Initiation Protocol - SIP [22] servers). Although *personal mobility* can be treated as a type of mobility, it is not directly concerned with the movements of hosts to different points-of-attachment in the Internet, but rather with the maintenance of updated pointers to a user's current location and identity. Another key feature of *personal mobility* is that it is only essential during the service initiation phase, that is, at the instant when a session is being initiated. When this session is initiated on the particular host to which a user is currently attached, the consecutive delivery of packets and managements of routing is mainly the responsibility of *host mobility*.

b) ***Host or Terminal;*** Internet hosts are identified by IP addresses. Maintaining the connectivity of hosts to any point-of-attachment in Internet networks is the task of ***host mobility***. In practice, this refers to the functions, which facilitate delivery of packets to and from the current point-of-attachment, or, in other terms, maintaining routing information[3] in the network. Hence, *host mobility* for IP addresses is a network layer issue (i.e. IP issues). Unlike *personal mobility* where the critical moment was the establishment of sessions through some prior registrations, *host mobility* keeps a continuous awareness of the network elements involved in distribution of packets during the lifetime of a session.

The topic of this document is mobility in network layer-centric environments thus promoting *host mobility* as the focus. This can be further justified if *personal mobility* is superficially considered to involve only higher layers such as the session or application layer but, as mentioned above, there is a dependency on the underlying network layer mechanisms such as the possibility of mapping identifiers to IP

[3] Note: There is a difference between the maintenance of routing information for host mobility and identifier-mappings for personal mobility. The former is the network layer data distributed in adequate routers while the latter is an upper-to-network layer transitional feature placed in dedicated databases in the Internet.

addresses. In the remaining parts of the document the term mobility only applies to *host mobility*.

2.2 Mobility Problem Statement

Application of location dependent IP unicast addresses for identifying hosts in Internet networks, despite various benefits, creates problems when used for identifying mobile hosts (MHs). Packets addressed to unicast IP addresses are always prefix-routed towards the network or one of its subnets, which share the same network prefix (netID) with the address. In the native IP setup a host can only receive packets in different parts of the Internet if it owns a topologically exact unicast IP address for each network visited. This shows that the basic setup of IP imposes restrictions on the potential movements of hosts. If no additional IP mobility support is available, a single-prefix addressed host can only achieve connectivity to the Internet by being constantly attached to the "home segment[4]" of the network. In this scenario, regardless of whether hosts are using the wired or wireless medium[5] to connect to the network, mobility is only possible at the link layer (layer two - L2). While this limited, layer two-constrained mobility scenario may be sufficient in some cases, the goal of efficient IP mobility in the Internet is essential due to the generic nature of IP as the global network layer solution (with the accompanying lower and higher layers) and the commercial expansion of IP-based telecommunication environments. It is therefore essential to provide mobility support in the IP layer so that MHs can achieve connectivity to the Internet from different networks and subnets[6] without requiring a pre-assigned IP address for each point-of-attachment.

---

[4] "Home segment" can be either a class A, B or C network identified by the netID (prefix) of the IP address or the home subnet (netID + subnetID). This is relevant to the administrative setup of the network. Regardless, it identifies the exact location of the IP address.

[5] Mobility is usually considered for hosts, which are using wireless technologies to connect to the network. This should not impose restrictions on hosts connected through wired medium although the limitations on physical movements are obvious.

[6] While a subnet can also be considered a network (or a part of it) the distinction between a network as a prefix- defined domain (possibly consisting of multiple subnets) and a subnet as a smaller part of the network seems appropriate.

The standard TCP/IP protocol suite supports changes of points-of-attachments across various networks and subnets by allowing **reattachments**[7]. Reattaching is achieved when a MH connects to a new subnet or network, and, provided there are no security constraints, configures its terminal by receiving a novel identity, i.e. a new IP address. This can either be achieved by manual configuration of the terminal or through the deployment of configuration protocols. The most popular host configuration protocol is the Dynamic Host Configuration Protocol (DHCP) [24]. Additionally there are other scenario-specific configuration protocols such as Bootstrap Protocol (BOOTP) [25], Reverse Address Resolution Protocol (RARP) [25] and Point-to-Point Protocol (PPP) [25]. Essentially, all configuration protocols maintain an address pool administered by a server-type device, which then allocates IP addresses to hosts attaching to the network visited. These solutions are far from being acceptable for the current demand for mobility due to two crucial shortcomings:

a) Delays incurred during the negotiation phases before the new address is obtained. This is further complicated if the mobile host needs to advertise its new address to potential initiators of sessions (eg. through DNS).
b) The inevitable break-up of established sessions each time a host changes subnets or networks and runs a configuration protocol. Since hosts acquire a new IP address upon every configuration, transport layers (eg. TCP) would naturally need to establish a new source-destination end flow.

While these two shortcomings may be acceptable for hosts which rarely change networks and when they do so they remain fairly static, for MHs which are frequently changing points-of-attachment, fast IP mobility support without significant disruptions of established sessions is essential (certainly, break-up of sessions is unacceptable). The primary goal of IP mobility support is to enable maintenance of established sessions with minimal disruptions while not restricting the movements of MHs to a single point-of-attachment, subnet or network. This can be solved by the deployment

[7] Sometimes performing reattachments is referred to as **portability**.

of IP mobility protocols. Generally, IP mobility protocols are "special case" routing protocols, which attempt to provide adaptable routing and address translation mechanisms. One of the most crucial parts of any IP mobility protocol is the method of maintaining connectivity of a MH to the Internet while it changes its points-of-attachment. This procedure is called **handover**.

There are two aspects of routing in mobility scenarios. The first aspect is the delivery of packets to MHs, referred to as **downlink** routing. The second aspect is the transmission and consecutive routing of packets from MHs to relevant destinations, referred to as **uplink** routing. Due to the nature of IP and accompanying protocols, downlink and uplink routing in mobility scenarios can be solved by different mechanisms. For downlink routing, delivery of packets to a new destination in a visited/foreign network[8] can only be achieved if there is a mechanisms for interpreting the correlation between the MH's home and new location so that packets can be re-routed to it. This problem of mapping home IP addresses of MHs to their current or future identities and then the consequent efficient routing, is the principal concern of downlink routing. Unlike downlink routing, uplink routing can be performed by standard IP routing since MHs can transmit packets to their recipients in the same manner as if they were in their home network since IP routers perform packet-forwarding decisions based mostly on the destination address of the packets. However, it should be noted that this is a pure mobility routing perspective on the uplink routing problem, which may turn out to be more complex in cases where there are security constraints imposed on packets routed from foreign networks. This is mainly because the source address of packets transmitted by MHs corresponds to the MHs' home addresses and in some networks, outgoing packets, which do not contain

[8] Foreign networks can be defined from different perspectives such as regarding foreign networks as ones that belong to different operators or administrative domains. However, for mobility investigation this is not the case and a foreign network is defined as the one for which the original identifying unicast IP address of the MH, referred to as mobile host's **home address**, does not correspond. This is because packets addressed to an MH's home address are always prefix-routed to its original subnet or network. Thus, in situations where a particular IP network (defined by the network part of the assigned address(es)) is divided up into several subnets, moving away from the original subnet to a subnet belonging to the same network still incurs mobility support and promotes the new subnet as a "foreign network".

the source address from that network, may be discarded by security agents called firewalls.

However, leaving aside the possible restriction on topologically incorrect source addresses and considering the mobility related issues only, it can be stated that IP mobility is mostly concerned with downlink routing where additional uplink routing support may be required in some specific situations (for example MH-to-MH routing where both hosts are in the same foreign network, see session 3.4.3.4). For uplink routing, it is assumed that MHs send packets to static hosts as their destinations. This may not always be the case and destinations can also be mobile thus complicating the packet delivery. It is assumed that this is not the issue for the uplink routing of a particular MH but it again concerns the downlink routing for the recipient. This assumption is further justified in the designs of mobility protocols which are explained in the remained of this document and comes from the initial step of the uplink communication where any host in the Internet initially attempts to contact its destination based on the home address of the destination host. Hence, the subsequent mapping of a destination's home and new identifiers is a concern for downlink routing of the recipient[9].

2.3 IP Mobility Protocols Essentials

There is a great variety of IP mobility protocols currently available in the Internet research community. In order to understand some of the key protocol mechanisms used in those proposals and to assist in extracting the common design goals for

[9] Although the statement for uplink routing is generally true, some exceptions may occur. For example, some Regional mobility protocols rely on specific setups for IP networks where that particular protocol is deployed hence adopting unique scenarios for both uplink and downlink routing. Cellular IP [35] (see section 2.3.3.2) assumes that the network, where mobility is solved, constructs the same set path consisted of routing entries for downlink traffic as the path taken for the uplink traffic to reach the global Internet. Thus, uplink traffic may appear to be managed by the network, but not conceptually, only because of the particular setup of the protocol. In Cellular IP, this is needed, not for the purpose of routing the uplink traffic, but because the uplink traffic assists in updating/refreshing the routing entries for the downlink traffic. Again, even in Cellular IP, the uplink traffic is still routed hop-by-hop (an analogy to shortest path routing for this particular network setup) toward the "outside", that is, the global Internet.

mobility protocols, it is necessary to briefly explain some essential protocols for solving mobility. The most important protocol is Mobile IP [28], which, due to some of its protocol features and its status in the Internet Engineering Task Force (IETF) as the ubiquitous mobility protocol, has become the reference point for most of the more recent attempts for solving IP mobility. Additionally, some features and concepts introduced in Mobile IP are reused in many other protocols in various manners. This is further elaborated in the remainder of the document.

2.3.1 Mobile IP's basic protocol mechanisms

Mobile IP has introduced the term *care-of-address*, a temporary identifier for MHs when they are not able to establish connectivity to the Internet using their home addresses. *Care-of-address* is obtained in a foreign network and it is a location dependent unicast IP address which points to the particular foreign network where it is allocated to the MH. The principal features of Mobile IP are the mapping between the home address of a MH and its temporary *care-of-address* and the consecutive delivery of incoming packets to the current location, that is, the *care-of-address*. A specific mobility agent called the **Home Agent** (HA) located in the home network of the MH performs the mapping. HA contains a form of routing entry for all MHs, which belong to its particular network segment and are currently achieving connectivity in foreign networks by the use of *care-of-addresses*. The key operation of a HA is the interception of packets addressed to "absent" MHs by sources (Corresponding Hosts (CH)) and then, look-up of routing entries to extract the current *care-of-address* for those hosts. The next step, after the mapping between the home address and the *care-of-address* has been resolved, is the delivery of packets to MH's current location by performing encapsulation [43][44][45] of the original packets into new packets addressed to the current *care-of-address*. Thus, an **IP tunnel** is created from the HA to

MH's current location. HAs can either be IP routers or hosts (even virtual home networks) with the essential ability to intercept packets addressed to MHs' home addresses for which routing entries exist. This feature can be achieved either by simple "snooping" of the packets (usually if the Home Agent is also an IP router for the subnet) or by other mechanism such as Address Resolution Protocol (ARP, proxy or gratuitous)[25].

In order to receive packets in foreign networks MHs need to perform several protocol steps for acquiring and then registering the *care-of-address* with their HA. There are two types of *care-of-addresses* distinguished by the way in which they are obtained (see Figure 2.1):

a) Foreign Agent *Care-of-address*: In this scenario MHs initially achieve connectivity with mobility agents called Foreign Agent (FA) which are located in foreign networks and are assumed to share the link with MHs. This is achieved via the Mobile Agent Discovery procedure, which is a mobility-extended procedure laid out in ICMP [11] for Router Discovery [26]. The procedure requires FAs to periodically transmit **Agent Advertisements** (mobility-extended ICMP Router Advertisement message) or alternatively to respond to **Agent Solicitation** messages sent by MHs. Agent Advertisements contain an address of the FA as a *care-of-address*(*es*) for a MH (A FA may have more addresses belonging to it). **The final result is that a MH can use an address of the FA as its *care-of-address*.**

b) Collocated *Care-of-address* (CCoA): MHs may obtain *care-of-addresses* by some address acquisition mechanisms such as DHCP[10]. In this case, the MH owns CCoA for the duration of its use. CCoA should have the same network prefix as the link (subnet) a MH is attached to. In this case the MH perform decapsulation of incoming packets.

[10] Note: Although DHCP is used, it does not promote this feature of Mobile IP as a **reattachment** (portability) scenario explained in section 2.2. This is simply a mechanism for obtaining a *care-of-address* and consecutive packet delivery is still dependent on

Regardless of the way in which the *care-of-address* is obtained, MHs need to register with their HA to create fresh routing entries. This is done through the exchange of registrations messages, the Registration Requests, sent by MHs and the Registration Replies, sent by HAs as acknowledgments to Registration Requests. The difference between FA *care-of-address* and CCoA is that, in the former case, the Registration Request is initially sent to the FA, which then relays it to the HA while in the latter case, MHs send it directly to their HAs (some deviations to this scenario are possible). During the exchange of these messages there are various options available for specifying the exact manner in which a Mobile IP connection may operate. Finally, HA encapsulates packets towards the tunnel exit points (either the FA or the actual MH (if CCoA is used)), where the final decapsulation is performed and the original packet is delivered to the MH.

The uplink transmission from MHs is assumed to be performed by standard IP routing.

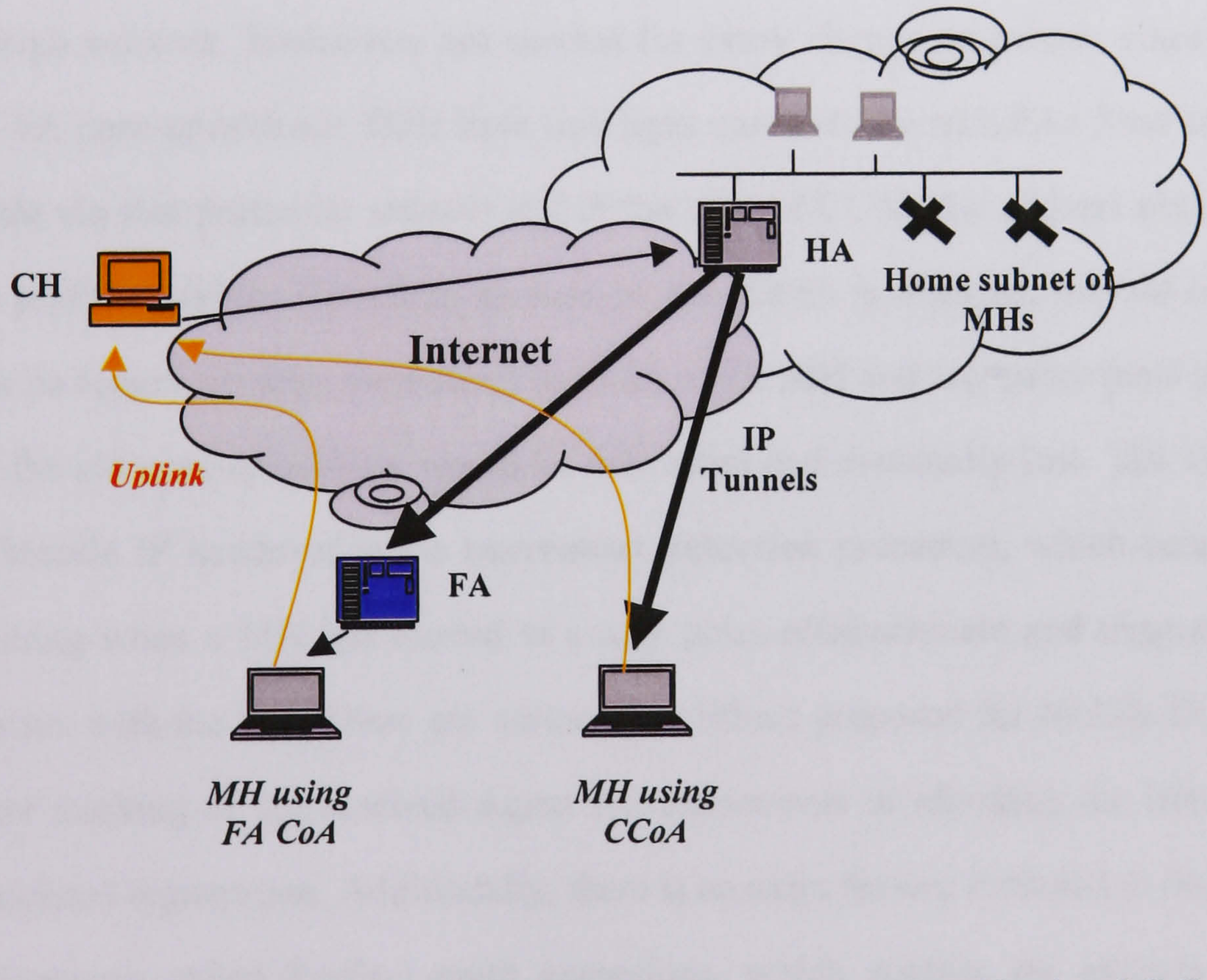


Figure 2.1. Two Mobile IP scenarios with MHs using FA *care-of-address* (CoA) and Collocated CoA (CCoA) options.

Mobile IP mechanisms. In practical terms, the Mobile IP protocol software still handles the incoming packet before it delivers them to other IP processes.

The base Mobile IP can be enhanced with a complementary protocol called **Route Optimisation** [27]. The main purpose of Route Optimisation is to overcome ***triangular routing***, a shortcoming of Mobile IP occurring in some scenarios when the direct route from a CH to a MH is shorter than the Mobile IP route taken to reach the MH from CH via the HA. Similar to the setup of Figure 2.1 where triangular routing is evident, if packets are sent to MHs directly from the CH hence avoiding the HA, they would experience less delay in reaching the hosts and *triangular routing* would be avoided. In Route Optimisation, CHs store routing entries for mobile hosts called ***bindings*** by contacting HAs and exchanging *binding messages*. Some other features are also possible such as the "smooth handovers" option for avoiding packet losses by having the old FA temporarily send packets to the new FA during handovers.

Handover in Mobile IP is achieved by performing re-registrations with the HA after obtaining the new *care-of-address*[11]. A MH, regardless of the type of *care-of-address* that it uses, essentially repeats all the steps performed during the initial connection to the foreign network. Handovers are needed for every change of subnet, since in the case of FA *care-of-address*, MHs have link layer connectivity with FAs (thus are only reachable via that particular subnet) and in the case of CCoA the address obtained is always prefix-routable. Therefore, as soon as the subnet is changed, the old *care-of-address* no longer presents the correct location of the MH and any subsequent packets sent to the old *care-of-address* would be misrouted and eventually lost. The key part of the Mobile IP handover is the **movement detection** procedure, which consists of determining when a MH has moved to a new point-of-attachment and triggering re-registration with the HA. There are various algorithms proposed for Mobile IP, based on either tracking of the received Agent Advertisements or checking the lifetime of the associated registration. Additionally, there is an extra feature included in the Agent Advertisements called Prefix-Length extensions, which contain the address prefix information for Foreign Agents so that consecutive Agent Advertisements can be

compared to check whether they belong to different subnets. In the case in which a different subnet is detected, based on the examined Prefix-Length, a handover is imminent. Generally, the movement detection works better for the FA *care-of-address*. This is partly due the presence of Agent Advertisements, which expedite the handover decision. In the CCoA case there are no Agent Advertisements involved and MHs are required to use other methods of detecting the surrounding subnet. Even if the subnet sensing is efficient a MH needs to search for an address distributing entity (such as a DHCP entity) thus further complicating the handover. Because of the difficulties related to the CCoA case, Mobile IP is naturally assumed to include the FA *care-of-address*. This is the approach adopted in the remainder of this document.

2.3.2 Design principles of IP mobility protocols

Historically, Mobile IP was the first significant effort for IP mobility support and has emerged as the starting point for the development of more efficient protocols. As was mentioned in the previous section, *triangular routing* in Mobile IP may incur packet delay depending on the positioning of mobility agents. *Triangular routing* does not present the most serious drawback of Mobile IP since it is relative to the positioning scenario and mostly affects the initial downlink flow of packets until they reach the MH. The consecutive delivery of packets, assuming the flow is consistent and there are no substantial delays or jitter experienced by packets, remains steady and indeed represents the transmission behaviour of the source. All packets in Internet communications experience a delay before they reach their destination. The difference in triangular routing is that there may be an additionally offset delay incurred due to the non-optimal route from a CH to a HA and then to a MH. Although this drawback needs to be tackled and a solution such as Route Optimisation is beneficial, it does not represent the essential drawback of Mobile IP. More importantly, the main

[11] Some differences may occur if the "smooth handover" feature of Router Optimisation is used.

shortcoming of Mobile IP is the ***handover latency*** when a MH changes a FA and attempts to re-register with its HA (the same delay is applicable to the CCoA case). There are two main network-layer-related causes for ***handover latency***:

a) *Handover Execution Delay*: This delay is caused by the movement detection procedure: the decision-making process during the initial phase of the handover and the consecutive re-connection to the new point-of-attachment. The complete analysis of the handover management is presented in Chapter 6 where the pros and cons of the movement detection procedure are detailed and alternative generic improvements assisted by a link-layer support are proposed. However, as far as the pure network layer procedure of the movement detection in Mobile IP, it does not present a significant drawback of Mobile IP since this type of procedure has to be performed (similar or identical) regardless of the mobility protocol deployed.

b) *Registration Delay*: After the completion of the re-connection to the new point-of-attachment, a new FA is selected and a Registration Request is sent to the HA. This step can induce significant delays depending on the distance between the HA and the MH's current point-of-attachment. The overall delay "absorbed" by the Registration Request is the summation of delays in the Internet along the route from the new point of attachment to the Home Agent: packet-processing, transmission, propagation and other forwarding-scenario based (congestion, routing policy…) delays. The resulting effect is usually the loss of packets, which are "blindly" sent to the old FA by the HA before the new Registration Request is received. For UDP sessions this creates a "hole" in the sequence of packets received and the inevitable slow down of the packet transfer rate for TCP, which in some cases, leads to severe degradation of transmission throughputs. *Registration Delay* is the most significant drawback of Mobile IP.

Reducing the handover latency (specifically the *registration delay*) is the main reason for the post-Mobile IP development of various mobility protocols. Mobile IP does provide an extension for overcoming this problem by allowing the existence of

simultaneous bindings in a HA for both the new and the old point-of-attachment (FA) so that packets can be delivered to both the old and the new FA. However this solution is far from being the ultimate answer for overcoming the *registration delay* due to the difficulties in its realisation: HA willingness to accept simultaneous bindings, ability of MHs to determine their next FA and update the HA accordingly, maintenance of simultaneous connection to old and new FA. Additionally, having the registration messages traverse across the Internet between the mobile host and its HA and instructing the HA to duplicate packets to both FAs, again, across the whole of the HA-to-FA path, substantially overloads global Internet resources.

Defining the requirements for mobility protocols is a subjective process since it may depend on the network policy and infrastructure of certain administrators or operators. It is therefore beneficial to define some common design goals for development of mobility protocols, which could then meet the requirements of all Internet networks. Some essential design goals for mobility protocols are:

- **Minimising *handover latency***: Considering the key factors influencing *handover latency* (presented in the above-mentioned Mobile IP case), all mobility protocols attempt to facilitate efficient and fast changes of points-of-attachment and hence to incur minimal transmission disruptions. As mentioned before, this is one of the driving reasons for development of efficient mobility protocols.
- **Reducing the protocol overhead**: Protocol overhead is the impact of the protocol mechanisms on the resources of a network. This includes bandwidth consumption in the network links involved due to transmission of protocol control messages and processing and memory use in the routers dealing with those protocol control messages. Protocol overhead needs to be reduced in all parts of the Internet involved in the communication: foreign network, global Internet[12] and the home network. The limit on the amount of resources available to MHs in foreign and

home networks is usually relative so the constraints on the protocol overhead may vary. However, the global Internet, under all circumstances, should be subjected to minimum possible protocol overhead.

- While satisfying the two efficiency requirements mentioned above, mobility protocols need to provide a solution with desirable characteristics. The term optimum may seem rather vague but experience in mobility indicates that a mobility protocol may actually achieve both satisfactory *handover latencies* and maintain an acceptable protocol overhead and still not emerge as an acceptable solution for wide deployment. A mobility protocol may be considered to possess optimum characteristics if it is: scalable (eg. performs well in different scenarios of populations, MHs and network size), robust (eg. reliable and adaptable to any failures), easy to deploy… Although sometimes subjective, this set of characteristics may ultimately affect the popularity of a certain mobility protocol.
- **Compatibility with other protocols**: Recent trends in Internet development indicate that many protocol designs, including the mobility ones, strive to harmonise their operation with other protocols in the Internet. Already there are proposals for dealing with integration of mobility and QoS protocols [3].

If mobility is observed from an abstract perspective there is a common set of functionalities arising from the nature of IP and mobility in Internet networks (section 2.2), which need to be dealt with by all mobility protocols. It is seems inevitable that a MH needs to obtain another identifier[13] while it is seeking connectivity outside its home location. Thus, a mapping between the MH's new and home addresses needs to be maintained. A concept of a **Location Directory** (LD) can be defined where updated information about the current location of MHs can be stored. The properties and locations of a LD are relative to the mobility protocol, the key operations being

[12] While all networks, both foreign and home, can be considered as Internet networks, applying the term "global Internet" introduces an illustratory scenario for mobility observation. Although generally assumed in mobility investigation, foreign and home networks do not have to be stub domains.

the maintenance of "routing-entry-like" mappings between the home and the new address of a MH and a constant refreshment of the mappings contained through a frequent and reliable communication between LDs and associated MHs. This communication usually comes in the form of updates sent by MHs. The most critical part of the communication arises during handovers because the new location needs to be updated in the corresponding LD as promptly as possible to avoid packet losses, which occur during periods when the LD contains the stale mapping.

Finally, for delivering packets from sources (CHs) to MHs, there needs to be a mechanism for **redirecting** packets from the original destination (the home address of the MH) to the MH's current location. LD plays a key role in this since it contains the address mapping for MHs. In this abstract mobility model, LDs do not have to perform the redirecting, which can be carried out by another entity provided it is constantly updated by the mapping contained in the LD. (in practical terms there are no restrictions, LDs and the redirecting points can actually be sub-elements of a single entity: an agent, a router or a distributed set of routing entries). Redirecting is usually achieved by **re-addressing**[14] the packet to the new location.

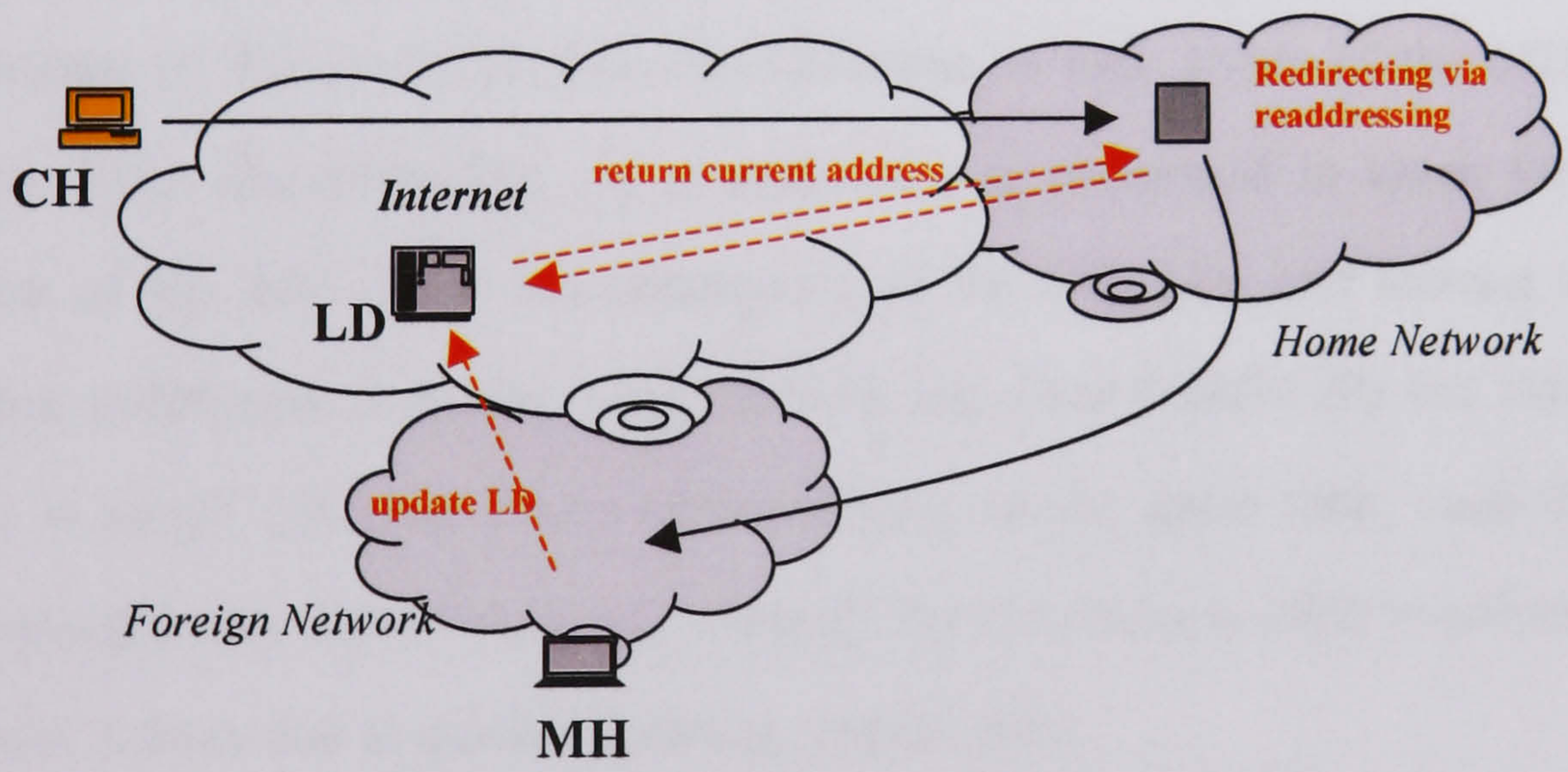


Figure 2.2. An example setup of the key processes in the abstract mobility model.

[13] Usually an identifier away from the home network means another IP address. However, this may not literally mean that a MH is always allocated a temporary IP address; it can actually use an address of a Mobility Agent (such as FA *care-of-address* in the Mobile IP case).

[14] At the initial glance this may seem inevitable, because to redirect packets they have be sent to the new location identified by the new IP address (*care-of-address*). However, as is evident from the following sections and chapters of this document, some LDs are implemented as a set of distributed routing entries where a change of IP address is not required for every change of point-of-attachment (micro mobility case).

An example of how the elements of the abstract mobility model may function is shown in Figure 2.2. The LD is placed in an arbitrary location in the Internet indicating that, regardless of the location, it needs to be refreshed with the MH's current location. Redirecting of packets is achieved in the home network by requesting addressing information from the LD.

Recollecting the Mobile IP case, it can be deduced that there is a straightforward similarity between the HA and the concept of a Location Directory and that the redirection and re-addressing is also performed in the Home Agent. It is evident that the main reason for *handover latency* is the delay incurred during the updating of the mappings held in LDs (in the Mobile IP case this was called the *registration delay*). The logical conclusion to this problem is to distribute the LD in such a way so that the updating of LD's mappings, which is usually a matter of sending a control message to the LD, takes as minimum a time as possible. This can only be achieved if LDs are placed (distributed) in the vicinity of MHs.

The nature of IP dictates that any new host wishing to send packets to a MH, at the first instant of the communication establishment, is only aware of the MH's home address. Then the redirecting via re-addressing is performed to reach the current location of the MH. From the perspective of the CH, LDs can interact and thus facilitate redirecting from the home network (eg. base Mobile IP) but can also be placed in actual CHs (eg. Route Optimisation). At the same time, once this **basic redirecting** from a CH is achieved, "moving" the LD closer to MHs would reduce the *handover latency* due to quicker updating (registration).

To achieve this goal the LD needs to be **distributed** from the HA or CH towards the foreign network where a MH is located. This represents the principal design objective of all mobility protocols.

The elements of the abstract mobility model are unavoidably present in all mobility protocols through deployment of the protocol's mechanism. The functionality of these

protocol mechanisms is one of the key differentiating factors used in classification of protocols.

2.3.3 Classification of Mobility Protocols

IP mobility protocols can be broadly divided into Global and Regional protocols [5] by considering the method of distribution of location information (i.e. LD) throughout the network. Mobile IP is the reference example of a Global mobility protocol. Generally, it keeps the location information in the HA (basic redirecting, see the previous section) and does not attempt to distribute the location information towards MHs. Considering Mobile IP as a purely Global mobility protocol, it can be stated that all Regional Mobility protocols perform a step further in the realisation of LDs, i.e. they all attempt to distribute the location information closer to MHs. Due to the great variety of protocol mechanisms used, Regional mobility protocols can be further classified.

Although the split into Global and Regional mobility protocols is a descriptive representation of the essential differences between the two sets of mobility protocols, it does not impose any operational independence between the two categories. It seems inevitable that all Regional mobility protocols need to achieve the basic redirecting step (eg. through a function in the home network of the MH) at least during the initial phase of the session. During this initial phase, Regional mobility protocols function as Global mobility protocols and later distribute their operation into other regions. Thus a mobility protocol can be a collection of Global and Regional mobility mechanisms each responsible for different parts of the mobility support. This integration of Global and Regional mobility protocols, usually Mobile IP and a local mobility protocol, is sometimes referred to as a **macro** and **micro** mobility split with a clearly defined operational transition point between the two parts of the protocol.

The two major categories of *Regional Mobility* protocols, distinguished by the method of realising the LD, are:

***a)* Proxy-Agent Architectures (PAA)**

***b)* Localised Enhanced-Routing Schemes (LERSs)**

Some of the key IP mobility protocols are shown in Figure 2.3 along with the categories they fall into and very roughly how they relate to each other[15].

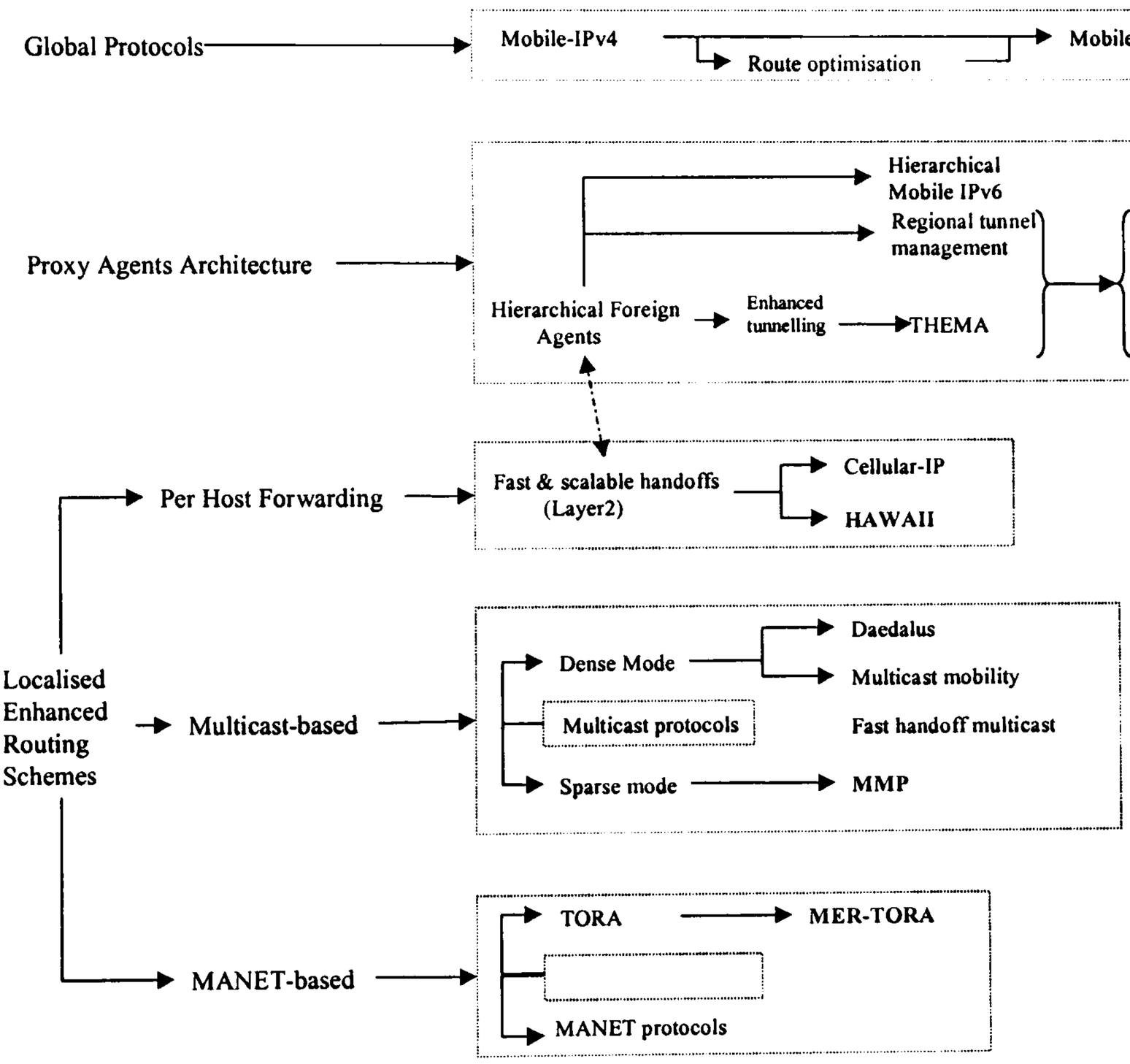


Figure 2.3. Classification of IP mobility protocols

[15] Note: While the claim here is that the classification indeed presents the division of mobility protocols along a well-defined axis, some similarity between different categories may exist and some protocols could include features of many categories. However, the primary concern of the classification of mobility protocol is not to provide a high level citation of all available protocols, but to assist in extracting the key mechanisms for solving mobility, which tend to be repeated (yet differently implemented) in protocols belonging to the same category.

2.3.3.1 Proxy Agents Architectures (PAAs)

PAAs realises the concept of the LD by establishing a hierarchy of Mobility Agents (see Figure 2.4), each containing a portion of the exact location information about a MH. These Mobility Agents are usually extensions of the FAs (rarely the HA) proposed in Mobile IP. Thus PAAs extend the basic idea of Mobile IP by introducing a layout of "FA alike" Proxy Agents, usually by placing them in foreign networks. Generally, packets are still traversing the HA, which sends them down the hierarchy of Proxy Agents toward the MH. This shows that PAAs, as Regional Mobility protocols, do resort to Mobile IP (Global mobility) for some aspects of the mobility support, as inicated in the previous section. An example operation of a PAA can consist of:

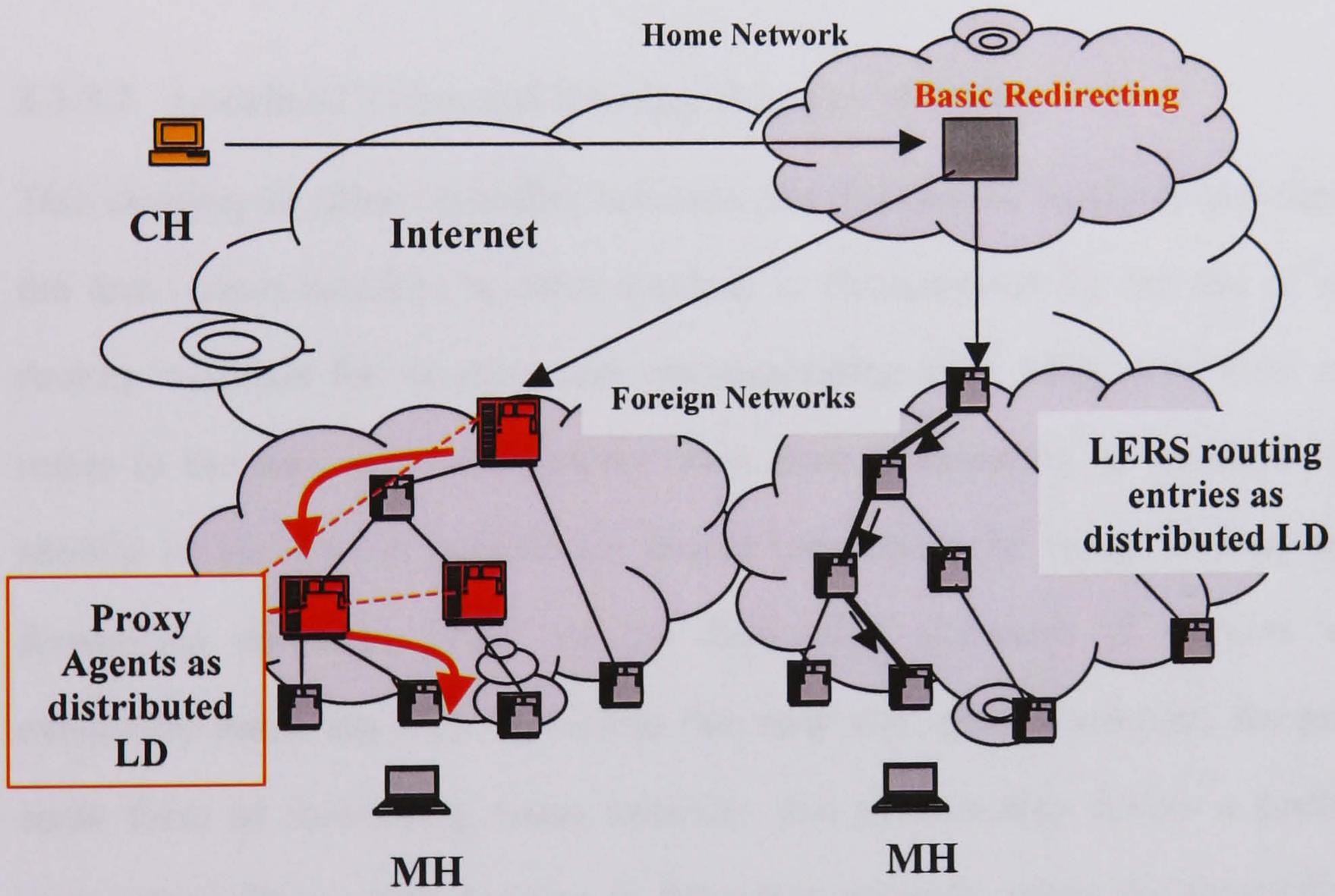


Figure 2.4. General mechanisms of PAAs (left) and LERSs (right)

- A MH registers with its local Proxy Agent ('a') at the bottom level of the hierarchy ("MH is at Care-of-Address"), which in turn registers with its nearest Proxy Agent at the next hierarchy-level ("MH is at Agent a"), and so on up the hierarchy

towards the HA. This way, when the MH changes its *care-of-address*, the registration request does not have to travel up to the HA but remains 'regionalised'. **This represents the way in which PAAs reduce the *handover latency* by localising the updating process due to distributed location information.** Packets from a CH travel down the hierarchy, usually by being tunnelled from one level to the next.

Examples of PAAs include the initial Hierarchical Mobile IP [29] and its alternatives, which place and interconnect Proxy Agents more efficiently: Mobile IP Regional Registration [30], Transparent Hierarchical Mobility Agents (THEMA) [31], Fast Handoff Methods [32], Hierarchical Mobility [4], Hierarchical Mobile IPv6[16] [33] and [4].

2.3.3.2 Localised Enhanced-Routing Schemes (LERSs)

This category of **micro mobility** schemes (for this type of Regional mobility protocol the term micro mobility is often applied) is characterised by the use of alternative routing methods for locating and communicating with MHs. The term alternative refers to the way in which packets reach their destinations. In standard IP routing, Mobile IP and PAAs, packets are routed (sometimes by using tunnels) by making forwarding decisions based on the destination addresses of packets until they eventually reach the MH. In LERSs this may also apply, however, the presence of some form of forwarding states indicates that packets may follow a predetermined route where the routing decision is formed in advance, when the forwarding entries were created[17] (see Figure 2.4). Just like the default requirement for any standards IP routing protocol, routers involved in packet communication should create forwarding entries, which point to recipients of packets. For LERSs, this means a type of entry for

[16] Note: There are IP version 6 protocols mentioned in this classification although this part of the document deals with mainly IP version 4 issues. All IP version 6 related issues are discussed in section 3.6.
[17] Obviously, the forwarding decisions still imply checking the destination address of packets to find out which actual entry corresponds to the packet.

each MH currently requesting services in the local domain. Thus in LERSs, the concept of the LD is applied in a distributed way by creating a collection of routing information in relevant parts of the network, i.e. in a scoped area of the foreign networks of MHs.

Fast and Scalable Internet Handoffs [34] is the initial effort resembling a *localised enhanced-routing* approach and proposes a solution for Local Area Networks (LANs) based on Ethernet, where gratuitous ARP messages are used for location updating, thus creating a forwarding entry in all nodes on the LAN and acting transparently to the rest of the Internet (actually Mobile IP was used again for macro mobility). Shortcomings of this protocol are immediately obvious due to the scope of implementation, which is restricted to LANs based on Ethernet.

There is an extensive variety of LERSs classified into three main categories:

- ***Per host Forwarding Schemes***: Inside a domain, a specialised path set-up protocol is used to install "soft state" host-specific forwarding entries for each MH inside a domain in a foreign network. The domain, which appears as a subnet to routers outside it, is connected to the Internet via a special gateway, which must be pointed to by the default gateway of the routers (or packet forwarding nodes) inside the domain. Examples include Cellular IP [35] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [36][37]. Cellular IP and HAWAII solve mobility via the deployment of new routing methods specifying the handling of control messages and router entries. Both protocols use the Gateway router as the interface between the Mobile IP-controlled *macro* mobility and *micro* mobility, and require participation of MHs in the setup and maintenance of forwarding entries. MHs have two identifiers: one *care-of address* for *macro* mobility (the Gateway's address) and an identifier in the *micro* mobility network section. Cellular IP simultaneously uses control messages and MH-originated packets as a triggering mechanism for establishing trees of forwarding entries, while HAWAII uses an explicit receive-and-acknowledge control procedure to set-up the routing trees.

While HAWAII researchers have introduced a mechanism for paging idle MHs [37], Cellular IP's initial design contained two types of forwarding entries: A routing and paging entry for active and idle hosts respectively having a smaller refresh frequency for Paging entries as the main distinction. Handovers are dealt with similarly by Cellular IP using timeouts to cancel out Routing entries in old base stations and HAWAII focuses on reducing the number of wasted packet on old links by redirecting their flow from old to new base station by using a redirect control message.

- ***Multicast-based Schemes*:** These schemes are explained in more details in the following chapter. Basically, multicast routing is used for creating forwarding entries and presents realisations of the concepts of LDs. Examples include *dense mode* multicast-based [38][39][40] and *sparse mode* multicast-based [1][2] mobility protocols.
- ***MANET-based Schemes***: Mobile Ad-hoc Networks (MANET) protocols were originally designed for ad-hoc networks, where all network elements are mobile, i.e. there is no fixed infrastructure. The routing is usually multi-hop and adapts as the MHs move and connectivity in the network changes. MANET protocols can be modified for IP mobility scenarios discussed in this document, where there is a fixed infrastructure and only hosts are mobile. Currently there is only one proposal in this category: MER-TORA [41] based on an ad-hoc routing protocol [42] for creating forwarding entries for MHs in the fixed part of the network (foreign domain). Thus ad-hoc protocols can be modified to perform LD functions on behalf of MHs attaching to the IP network.

CHAPTER THREE

Multicast for Mobility Protocol

Chapter Overview

This chapter details analytical steps that preceded creation of Multicast for Mobility Protocol – MMP along with the full description of the resulting protocol mechanisms of MMP. Focus is initially placed on explaining the analytical processes, which are the foundation for choosing multicast routing as the routing solution for solving mobility of IP hosts. This includes the abstraction of functional similarity of multicast and mobility and overview of relevant solutions that are used as the starting point for considering mobility with multicast and designing MMP. The main conclusion of the research in multicast as a mobility solution is presented as the three possible models of the integration where one of the models named Multicast-terminated Mobile IP is chosen as the most appropriate one and used as the functional foundation for specific protocol features of MMP. MMP is firstly described as an IPv4 protocol but the chapter then introduces features of IPv6 and describes the adaptation of MMP for IPv6.

3.1 Why Multicast for Mobility

Prior to considering the design processes involved in devising models and solutions for adapting multicast for solving mobility some general trade-offs can be observed. If a simple case is considered where IP multicast is used to reach Internet hosts, which are also mobile (i.e. MHs), there are the following **advantages** of applied end-to-end multicast routing (CH-to-MH):

a) *Transparent Mobility*: Multicast addresses as identifiers for MHs, i.e. multicast CoAs, eliminate the need for global updating during movements of MHs. Additionally, multicast routes are locally re-adjusted using in-built multicast mechanisms during mobility without the need for specific mobility-incurred end-to-end signalling.

b) *Automatic Handover Support Mechanisms*: Multicast routing is created to allow distribution of packets to any location in the Internet provided that the multicast route is appropriately formed. During mobility of MHs, redirecting of packets to new points-of-attachment is performed by the multicast tree forming procedures. Assuming that the packet transfer to the previous (old) point-of-attachment is facilitated with a multicast tree, moving to a neighbouring point-of-attachment incurs local updating, thus fast handovers.

c) *Ease of Deployment*: Applying the existing multicast routing for solving mobility in the Internet eliminates the need for designing new mobility mechanisms as the support can be provided via the exiting multicast routing protocols.

d) *Interactions with other protocols*: Multicast is already being considered in a variety of connectivity scenarios in the Internet. An example is the support for QoS protocols such as RSVP, which is naturally suited to interoperate with multicast routing.

At the same time using multicast routing for mobility involves the following **disadvantages**:

a) *Protocol Overhead Risks*: Internet-wide management of multicast addresses for allocation to MHs can be complex to manage with difficulties in achieving accurate and real-time states in address databases (e.g. DNS,...). Additionally, there are risks of multicast routing overhead for long–haul end-to-end multicast support, flooding risks (property of some *dense* multicast routing protocols), signalling overhead...

b) *Lack of the current Internet support for IP multicast*: Although multicast routing protocol discussed in section 1.4 are functionally complete and are ready for implementation in the Internet environments, only few of them are available in some parts of the Internet. From the perspective of using multicast for solving mobility, this lack of Internet support for multicast can be seen as inducing extra complexity since routers are not fully equipped with multicast capabilities, i.e. multicast routing protocols. In cases where the IP multicast supports are partially available, direct routing paths for end-to-end packet transfer may not be currently possible.

c) *Requirements for Modification of the Standard IP functionality*: Depending on the setup of the end-to-end communication, using IP multicast for supporting mobility can require extra underlying support in CHs, TCP implementations and ICMP.

The following text describes the technical issues for integrating multicast and mobility with a primary objective to benefit from the advantages and overcome the disadvantages of the general application of multicast for solving mobility shown above.

Internet multicast and mobility protocols share some common design goals, although these two sets of protocols are intended to solve entirely different issues in Internet communications. Recollecting the abstract mobility model presented in section 2.3.2

and the principles of IP multicast outlined in section 1.4 some general functional commonalities can be extracted:

a) ***Abstraction of the location independent addressing***: IP mobility protocols achieve this by performing re-addressing based on the stored location information about current addresses of MHs. The entire operation of a mobility protocol can be summarised as an attempt to allow the Internet hosts to remain reachable at any location in the Internet, thus reducing the importance of topological locations of hosts. Hence, all protocol operations related to re-addressing (eg. home to *care-of-address* transition) can be generalised as an abstraction of the location independent addressing facilitated by mobility protocols. With IP multicast protocols, the generalisation is more obvious as the location independent multicast IP addresses eliminate the requirement to maintain geographical information about the recipients of packets. This is supported through the execution of multicast routing protocols.

b) ***Efficient packet routing***: Both IP mobility and IP multicast protocols use the associated routing mechanisms to create particular routing entries in the networks involved. These entries then facilitate packet forwarding towards the end host(s). The general requirement for efficient packet forwarding and optimal maintenance of routing tables applies to both sets of protocols: looping freedom, minimised protocol overhead, interactions with the underlying IP routing protocols, need for integration with other protocols…

c) ***Adaptable location management***: Regardless of whether an already active multicast group member is changing its location with respect to the established multicast routing tree or a new member is joining a multicast group and awaiting the creation of the new multicast routing tree branch, all multicast routing protocols aim to dynamically adapt the routing (hence the location) information to different host behaviour patterns. The same requirement is present in the design of IP mobility protocols. This is highlighted during handovers, when MHs change

points-of-attachment and the mobility protocol readjusts the routing information (i.e. location information) according to new positions of MHs.

These functional similarities were the foundation for the early development of multicast-enhanced mobility protocols classified as *Multicast-based schemes* in the previous chapter (see section 2.3.3.2). The two pioneering examples of the use of multicast mechanisms for solving mobility are Mobility Support using Multicasting in IP (MSM-IP) [39] and the Daedalus protocol [38].

3.1.1 Overview and Analysis of relevant solutions[1]

The Daedalus protocol utilises multicast routing for packet delivery from HAs to MHs and uses Mobile IP for the global routing from CHs to HAs. The *care-of-address* stored in the HA is a multicast address. Since the HA sends packet to MHs using this multicast *care-of-address*, it requires a multicast routing protocol for delivering the packets to MHs. In fact, the protocol delivers the packets to Base Stations (BSs) (BSs are IP multicast capable routers with wireless interfaces and a wired one attached to a wired LAN - in this particular example, to an Ethernet network. DVMPR multicast protocol is used for multicast routing from HAs to BSs. MHs perform specific registration functions and do not actually join the multicast group. They instruct BSs to join the group routed at the HA for consequent packet delivery. Original packets are encapsulated in multicast packets by the HA. BSs are required to perform the final decapsulation and delivery to MHs. The main reason why IP multicast is deployed in this proposal is the ability of a MH to connect to more than one BS simultaneously by having the neighbouring BSs join the multicast group on behalf of the MH. A MH achieves this by sending registration messages to all neighbouring BSs. Only one BS

[1] The layout of this chapter suggests that the two analyzed protocols (Daedalus and MSM-IP) present earlier attempts of adapting multicast for mobility. This is in fact, adopted throughout the document. Actually, development of MMP somewhat coincided with the development of the two protocols. Since

is selected as the serving BS and requested to forward packets to the MH, while the others are instructed to buffer received packets without wireless forwarding.

During handovers, MH and BSs are not required to re-register with the HA since the new BS is already receiving packets because it has previously joined the multicast routing tree for the MH's unique multicast group. The only handover operation is the instruction sent to the new BS to stop buffering and start forwarding packets to the MH. This protocol is tested on a small-scale test-bed, which is actually an Ethernet link shared by both BSs and HA. A variety of functional additions are proposed at the relevant routing entities such as HAs, BSs and MHs to facilitate the transition between the Mobile IP and the multicasting part of the protocol and to provide necessary protocol steps for MHs in the wireless medium. Performance results, although highly dependent on the setup of the test bed, succeeded significantly in achieving the target *handover latency*[2].

Apart from presenting the first multicast integration in a mobility support system, the Daedalus proposal identifies some specific implementation mechanisms such as a set of changes to routing modules, BS's beaconing model and the buffering solution.

MSM-IP deploys a full scale, end-to-end multicast for mobility support for every MH. This eliminates the need for Mobile IP as the global routing mechanism and requires sources of packets to send them directly as multicast packets. MSM-IP provides an extensive insight into the problems and solutions related to this type of multicast deployment. Sources are assumed to perform the standard look-up of distributed session directories for discovering multicast *care-of-addresses* of MHs and then forward the packets to their local multicast routers. This step stands for the general redirecting phase in the abstract mobility model (see section 2.3.2). From this point

---

this thesis argues that MMP presents a more evolved approach to the problem, the chronological ordering of protocols is used.

[2] For the Daedalus proposal, handover latency is defined as the time between the reception of the first packet from the new BS and transmission of the handover-invoked forwarding request to the BS.

onwards, the local multicast router should run any available multicast routing protocols to reach the MH.

Although MSM-IP is flexible and can be used with both the *dense* and *sparse* mode multicast routing protocols, the protocol extensions presented mostly relate to *dense* mode environments and in this case the preference is given to DVMRP. MSM-IP does not use the standard "broadcast-and-prune" procedures for creating routing trees to MHs. Instead MSM-IP proposes a location management procedure using hierarchical scoping of assigned multicast addresses for discovering location servers of particular MHs. These location servers contain the address of the local multicast router of a MH so that the local multicast router of the CH can send (encapsulate) packets to the MH's local router directly without the overhead of the standard multicast tree-building procedures. This procedure, along with the session directory look-up, represents an actual realisation of LD (see section 2.3.2). Similar to the Daedalus proposal, MSM-IP benefits from the routing flexibility of multicast and allows neighbouring BSs to join the same multicast tree as the MH, hence achieving smooth handovers. Joining procedure is MH-controlled through the use of IGMP. Test bed deployment of MSM-IP shows that supporting protocols, such as TCP, ICMP and ARP, need to be modified to support full scale multicast for solving mobility. These problems mainly occur because of the application of the multicast address as the identifier (*care-of-address*) for MHs. One of the proposed solutions, requiring small modifications in the global network architecture, is to assign a temporary unicast address (not a CoA) to each MH as a network management remedy.

3.2 Models for integration of multicast and mobility

Besides the general functional commonalities of IP mobility and IP multicast presented in the previous section, one of the main practical reasons why multicast

routing is an attractive supplement to mobility protocols is the possibility of having multiple routing tree branches to current and anticipated points-of-attachment (i.e. BSs) of MHs. This is a highly efficient characteristic, especially with respect to *handover latency*, since it generally avoids delay-inducing registrations during handovers. Both Daedalus and MSM-IP proposals shows that this feature can be achieved regardless of whether MHs are the sole multicast group members for their "*care-of-address group*" (MSM-IP case) or they request the serving and the arbitrary neighbouring BSs to join the multicast group on their behalf (Daedalus case). The multiple multicast routing tree branches of *Multicast-based schemes* result in a very similar layout of routing entries (i.e. trees) as achieved in *Per-host Forwarding schemes*. Both schemes can also provide dynamic routing extensions for simultaneous registrations to more than one BS. The key difference between the two schemes is that the *Per-Host Forwarding schemes* propose new routing methods for handling multiple point-of-attachments while the *Multicast-based schemes* rely on the default property of IP multicast, which allows multiple tree branches for connecting multicast group members. The difference here is that there is only one "virtual group member": the MH, which requests multiple connections from neighbouring BSs.

Multicast can be adopted for mobility in various manners based on the scope of deployment of multicast in the whole mobility protocol. Scoping of multicast is related to its potential integration with global routing mechanisms, that is, Mobile IP. Hence three distinct model solutions can be extracted, as follows.

3.2.1 Full-scale multicast for mobility

The concept of this solution is adopted and implemented in MSM-IP taking into account the scale of multicast deployed. Multicast is deployed over the whole of the mobility protocol, from CHs to MHs. Multicast handles both the global and regional

mobility. CHs send multicast packets to MHs directly. Problems related to this solution are mainly associated with the scalability of multicast. Multicast addresses are used globally, hence introducing a risk of address exhaustion since there is no address scoping solution. Location management (LD is either completely or partially distributed throughout the Internet) is distributed globally because the MH's multicast *care-of-address* needs to be available for all potential CHs for initiating sessions (this then requires a type of LD solution proposed in MSM-IP: updating distributed session directories and/or implementing a look-up procedure for location servers). Thus, if a MH obtains a multicast *care-of-address*, that address becomes the "global identifier" for the MH and cannot be reused for any other purposes, either for another MH running a *full-scale* multicast for mobility solution (even the *hybrid* case, see next section) or for native multicast sessions for an arbitrary number of group members. Additionally, multicast has to be supported globally to create a multicast tree from the CH to the MH. This assumes that there is an "active population" of multicast routers from a CH to a MH. This can potentially induce significant overheads during the multicast tree establishment procedures (especially if a "broadcast-and-prune" protocol such as DVMRP is used to construct the multicast tree) or during the location management procedure for enabling the "virtual link-up" of multicast peer routers via IP tunnels (as proposed in MSM-IP). The following steps may be executed in a *Full-scale multicast for mobility* (see Figure 3.1):

- MH obtains a multicast *care-of-address* in a foreign network.
- CH learns about the multicast *care-of-address* of the MH, which is somehow globally circulated.
- A multicast tree is formed from the CH to the MH. This step may require a location management procedure similar to the MSM-IP proposal to avoid "blind" multicast tree establishment.
- CH sends data with **destination address = multicast *care-of-address***.

- Multicast routing protocol delivers packets from CH to MH. Depending on the availability of multicast routers in the path between the CH and MH, IP tunnels may be created between the routers.

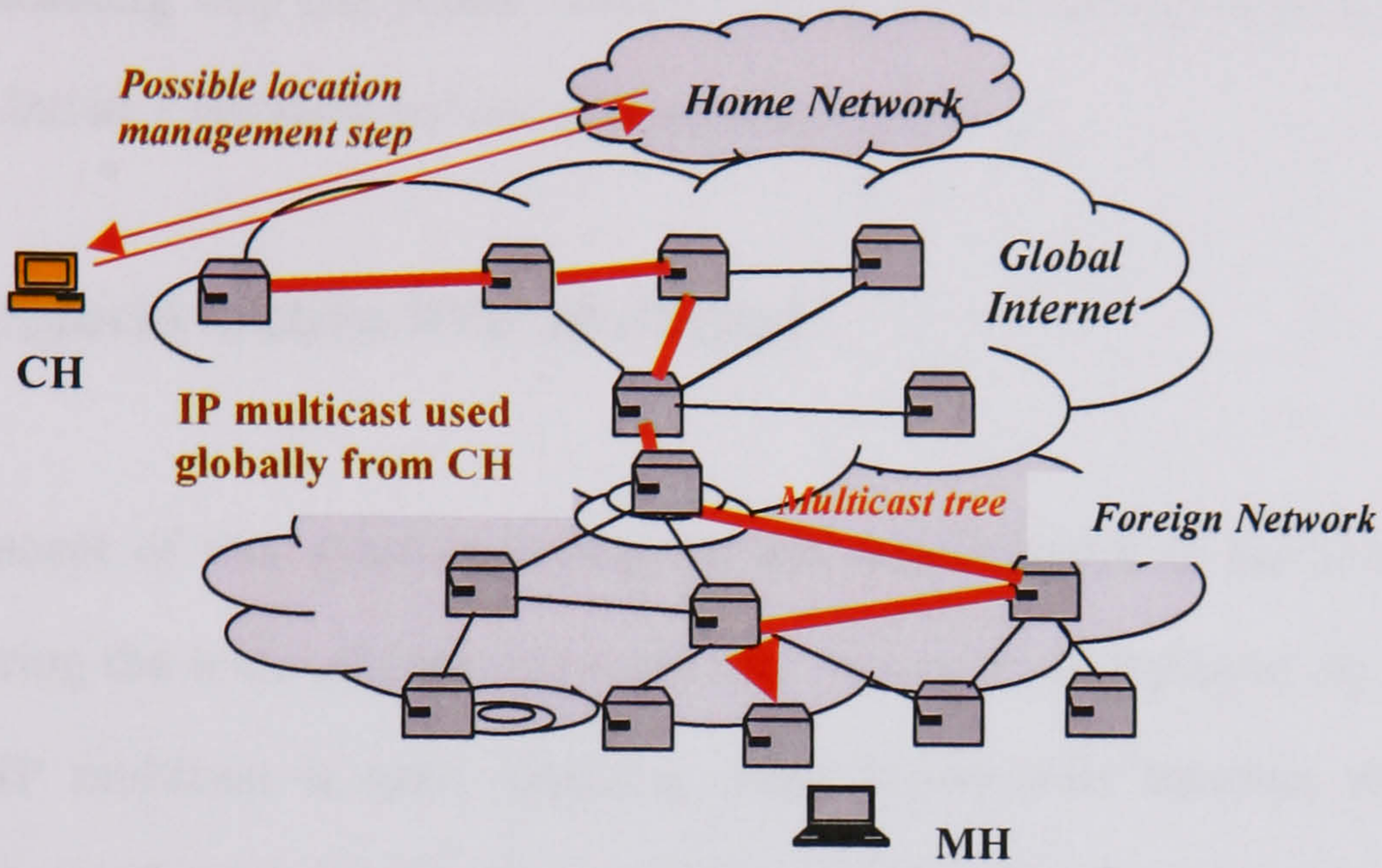


Figure 3.1. An example setup of the *Full-scale multicast for mobility* solution.

The explained example of *full-scale* multicast for mobility assumes that sources (CHs) send multicast packets directly without encapsulation because the MH uses multicast *care-of-address* in a collocated way, i.e. there is no home address to a multicast *care-of-address* transition and the multicast address is the unique identifier for the MH while it is in the foreign network[3]. Potential IP tunnels between multicast routers are of a different nature and are required for packet delivery between the routers after which the packets are decapsulated into their original packets addressed to the multicast *care-of-address*. A different scenario is possible where the CH (or the local multicast router of the CH) can encapsulate the original packets (addressed to MH's home address) into packets addressed to the multicast *care-of-address* of a MH. The communication would then experience the same protocol behaviour as the non-encapsulating case apart from the last step in the packet delivery because the packets would need to be decapsulated back again to reveal the original packets addressed to

[3] Sessions advertisement is required.

the MH's home address. Decapsulation would naturally be performed by the MH or a local router[4]. If the decapsulation is performed by the local router then the last communication phase between the local router and the MH would not be done in the IP multicasting way (no IGMP features are involved between the MH and the local router), but as a standard unicast delivery (eg. ARP...).

3.2.2 Hybrid Mobile IP/IP Multicast

The concept of this solution is adopted and implemented in the Daedalus protocol considering the scale of deployed multicast. Mobile IP is deployed up to the HA from where IP multicast is used requiring some interactions between the HA and the multicast routing protocol[5]. Thus, global mobility is solved by partially deploying Mobile IP and IP multicast. CHs send regular packets to the MH's home network as specified by Mobile IP. HA receives the packets, encapsulates and transmits them as multicast packets, which traverse the global Internet and the foreign network until they reach the MH. At the end-point, packets can be decapsulated by a local IP router (usually a wireless IP router, i.e. a BS if it has wireless connectivity with the MH) or the actual MH. If MHs performs the decapsulation then multicasting is implemented "all the way" from HA to MH and the MH can register using IGMP. Otherwise, as in the alternative *full-scale* scenario (end of the previous section), if the local router performs decapsulation, the last phase of the communication is done in a non-multicast manner (Daedalus protocol).

This model has similar scalability problems as the *full-scale multicast* case since multicast routing traverses the global Internet. In fact, the Mobile IP portion of the protocols can be seen as a location management alternative to the full-scale multicast

---

[4] Presumably, the local router is on the same subnet.
[5] In fact, HA is an IP router supporting IP multicast.

case proposed in MSM-IP[6]. Instead of running the location management procedure to obtain the multicast *care-of-address*, CHs normally send packets to the HA from where they proceed using multicast. The most significant difference between the two models is that the *hybrid* case always performs packet encapsulation from the HAs[7] whereas, in the *full-scale* case, the MH can be a "virtual multicast group member" and packets need not be encapsulated. The following steps may be performed in the *Hybrid Mobile IP/IP multicast* case (see Figure 3.2):

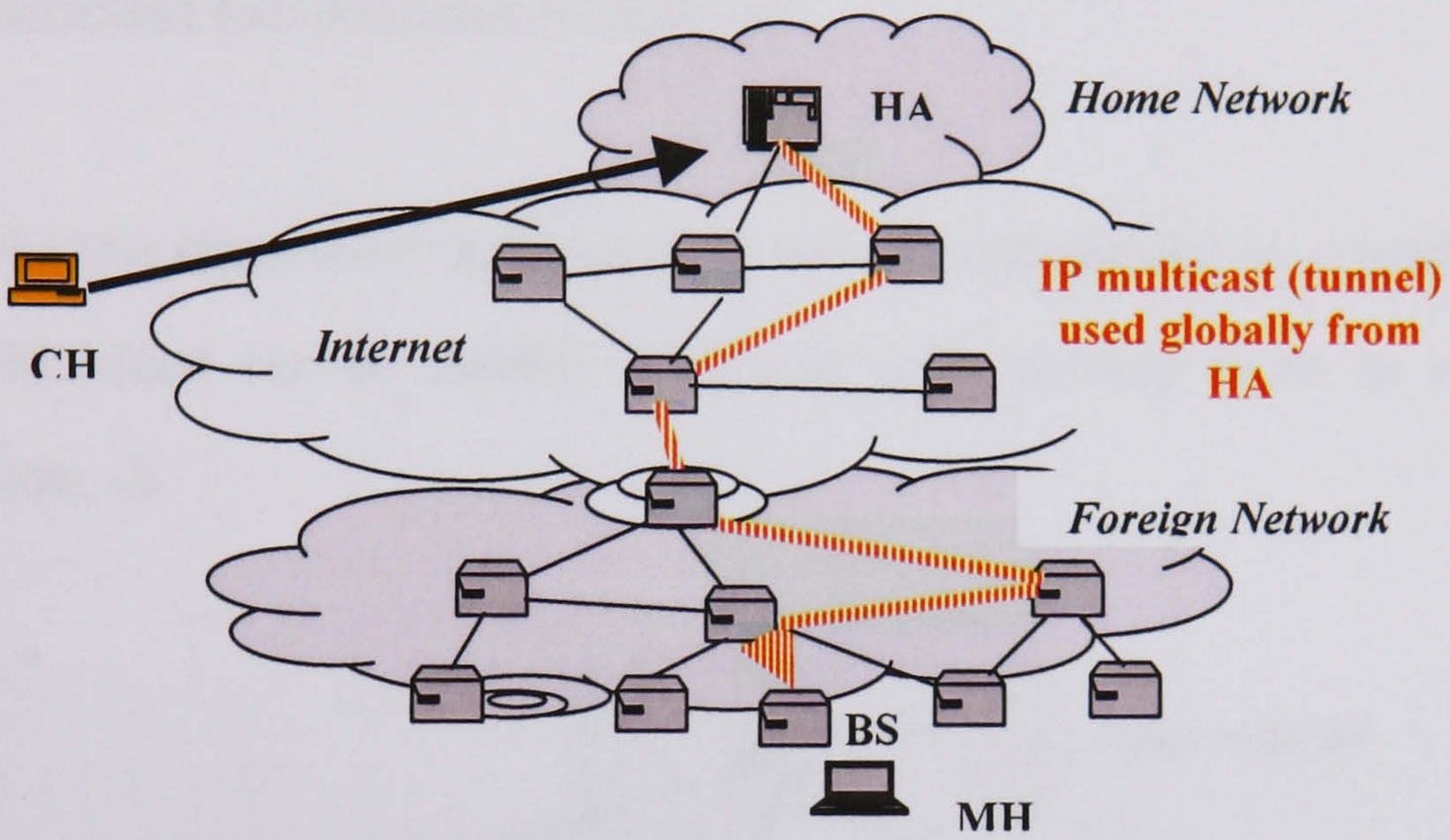


Figure 3.2. An example setup of the *Hybrid Mobile IP/IP Multicast* concept.

- MH registers and acquires a multicast *care-of-address* in a foreign network (either through pure IP multicast (IGMP) or by instructing the BSs to perform the multicast registration on MH's behalf).
- MHs or BSs (Daedalus protocol) update HAs with the current *care-of-addresses* (messaging can by done through Mobile IP control messages).
- CH sends packets with **destination address = MH's home address**.

---

[6] A Route Optimisation step where HA informs the CH about the multicast *care-of-address* would turn this *Hybrid* case into the *Full-scale* one with tunnelling from CH, provided there is multicasting support.

[7] Eliminating the tunneling from HAs in the *Hybrid* case would complicate the end-to-end transport layer association between the CH and MH and require extra transitional features in HA

- HA intercepts the packets, encapsulates them in the multicast *care-of-address* and forwards them using the multicast routing protocol all the way to the MH's BS.
- BS is a part of the multicast routing tree; it decapsulates and delivers the packets to the MH (either through the native multicast procedure or as a unicast delivery)

3.2.3 Multicast terminated Mobile IP

The two previous cases show that Mobile IP and IP multicast can be scoped in various ways. IP Multicast can be flexibly integrated with mobility (with or without the encapsulation…).

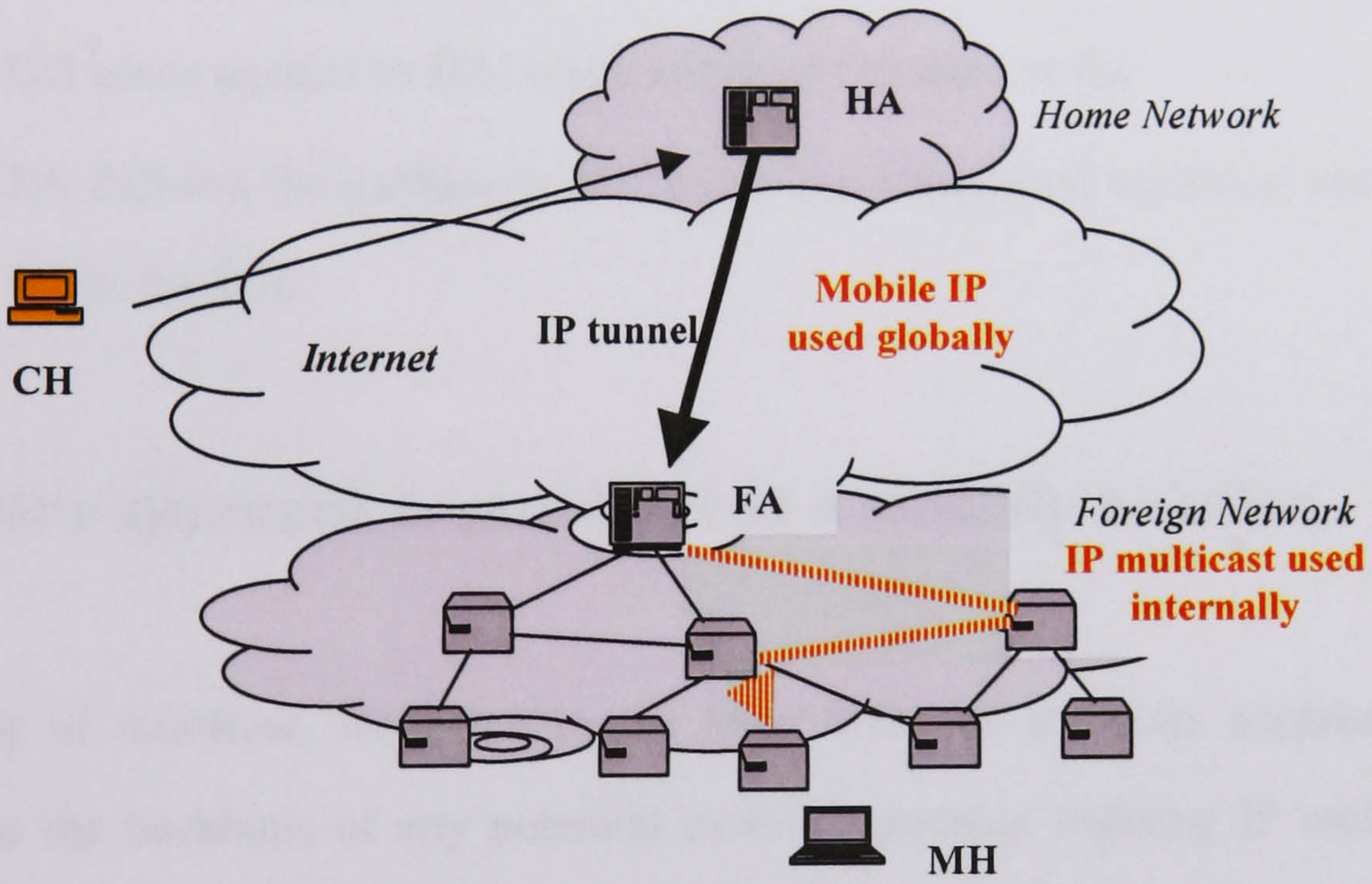


Figure 3.3. An example setup of the *Multicast terminated Mobile IP* concept.

In order to hide the application of multicast to the rest of the Internet, it should be deployed "below" the FA (or a router/gateway conceptually performing similar functions with respect to Mobile IP in the global mobility part), scoped in the regional mobility part of the whole protocol in an arbitrarily defined foreign network domain

(network domain is a scoped and uniquely administrated area possibly including different subnets and prefix-separated networks or merely a single link). Mobile IP is entirely used from CH via HA to a FA located at the ingress point of the foreign domain. Multicasting is deployed in the domain between the FA and the MH, thus using network layer multicast routing for handling movements of MHs inside the domain. Multicasting is transparent to the entities outside the foreign domain (CHs and HAs) and modifications are required in the FA to "transfer" the packets to the multicasting section of the protocols. The following steps may be performed if *Multicast terminated Mobile IP* is used (see Figure 3.3):

- MH registers with the gateway router (FA, in the Mobile IP part of the protocol).
- HA is only informed about the address of the FA, which it assumes is the address of the MH's.
- CH sends packets to HA, which encapsulates them to FA
- FA delivers the packets to MH along the established multicast tree from the FA to the MH.

3.3 New approach to multicast as a mobility solution

Scoping of multicast, identified in the three types of solutions explained above, presents the backbone of any potential mobility protocol utilising IP multicast. An important property of *full-scale multicast for mobility* is the requirement for a location management procedure for solving global routing issues. This is an essential requirement because in the absence of the location management step, CHs would send multicast packets to local multicast routers from where the process of reaching the current network of a MH could incur large delays and protocol overhead. The location management procedure (if the MSM-IP case is considered) essentially performs the

same functions existent in the global mobility routing by requesting the location of the MH's current network from a constantly refreshed server on the home network of the MH's. This conceptually mimics the operations laid out in the abstract mobility model (section 2.3.4) at least for the global mobility part. Accordingly, *Hybrid Mobile IP/IP multicast* can be regarded as the *full-scale* solution with Mobile IP as the location management extension. If the *full-scale* adaptation of multicast is considered, the best deployment scenario would actually be the *hybrid* case since Mobile IP is the standard solution for global mobility and its application would eliminate the need for a new location management procedure. When comparing the *hybrid* case as the main candidate of the *full-scale* solutions with *Multicast terminated Mobile IP*, the first observation is that the *handover latencies* in both cases would be roughly the same. This comes from the fact that multicast is responsible for tree reconfigurations during handovers for both scenarios[8] by forming routing tree branches to new point-of-attachments.

Multicast terminated Mobile IP is the most suitable mobility solution utilising multicast due to three advantages over the *hybrid/full-scale* solutions:

- **Reduction in the protocol overhead in the global Internet:** Protocol overhead is reduced because Mobile IP is used in the global Internet up to the foreign network domain, hence avoiding the tree-forming and packet-forwarding complexity of using IP multicast globally.
- **Robustness:** *Multicast terminated Mobile IP* is a robust protocol concept, as it deploys multicast transparently, relies on Mobile IP for global routing, which itself is a robust solution, and allows flexible and operator-dependent deployment of multicast in the foreign network domain.

[8] One exception to this may be during the changes of network domains when, in the case of *Multicast terminated Mobile IP*, routing would be handled by non-multicast mechanisms, that is, Mobile IP. The assumption is that this scenario can be omitted from the general discussion since as far as the mobility protocols are concerned; the primary concern is to provide efficient mobility inside the foreign network domains.

- **Ease of deployment:** Finally, deployment of *Multicast terminated Mobile IP* is significantly simplified by the transparency of multicast, which is dealt with by local mechanisms in scoped network domains. Thus, the protocol appears generic and can be used as standard Mobile IP if the particular network operator does not provide any additional routing supplements or does not want to deploy multicast. The split between the Mobile IP and the multicast parts of the protocol, clearly follows the same logic of the *macro/micro* mobility solutions where there is a split between the global mobility handled by Mobile IP and regional mobility handled by solution-specific protocol mechanisms.

The concept of *Multicast terminated Mobile IP* has been applied in the design of Multicast for Mobility Protocol[9] (MMP). Another important design decision during the development of multicast-based mobility protocols is the choice of the multicast routing protocol. The available examples of multicast-based mobility protocols claim operations independent on the choice of multicast protocol. However, both the MSR-IP proposal and the Daedalus protocol use DVMRP for the multicasting part of the solution. From the three examples of scoping of multicast for mobility (see previous sections) there is no direct correlation between the solutions and the type of multicast protocols, which could be incorporated. Additionally, there are no restrictions on the category of multicast protocol deployed: whether they are *dense* or *sparse* mode protocols, although the solutions available solely utilise *dense* mode protocols.

MMP uses a novel approach by deploying *sparse* mode multicast routing protocols instead of the conventionally used *dense* mode protocol. The endorsement of *sparse* mode multicasting is considered a significant design decision in making MMP a scalable, efficient and feasible mechanism for IP mobility. The main difference between *sparse* and *dense* mode multicast protocols (see section 1.4.2) is that *dense* mode protocols use variations of broadcasting (flooding) to distribute packets to

[9] MMP was designed in a British Telecom sponsored project. The intellectual property is shared between King's College London and British Telecom.

interested members/hosts while *sparse* mode protocols use a central distribution point, a router called Core (CBT) or Rendezvous Point (PIM-SM, shared tree part), to which sources send packets and interested members/hosts explicitly join. The explicit joining of the Core means that a host joins a multicast group by sending a control messages to the Core hence creating a routing tree to and from the Core through which packets can flow. A *dense* mode analogy would be very different: sources would broadcast packets in the whole Internet, until the interested member is reached, wasting significant bandwidth in the process. Routing trees of multicast routing protocols can be distinguished as source-based trees (separate tree for each source-to-receiver pair) for *dense* mode protocols and shared trees for *sparse* protocols. As the name implies, and the protocol mechanisms confirm, *sparse* mode protocols are more suitable for sparsely populated groups. Considering this fact further, because multicast in this case is intended for individual MHs and not a randomly numbered group, the choice of *sparse* mode protocols is more natural. The *sparse* mode multicast protocol chosen as the *micro*[10] mobility supplement protocol in MMP is CBT because it provides efficient, simple and fast tree forming and maintenance methods. PIM-SM is not chosen due to the redundant option of switching to source-specific trees and a more complex mechanism of tree establishment than CBT.

Figure 3.4 shows a **pure multicast** (not used for mobility but for a multicast group) scenario of an existing group with two new members joining the multicast tree with CBT Join Requests and CBT Join Acks building, first *transient*, and then *permanent* router states <*, group>. According to CBT specification, a Join Request does not always have to reach the Core but it can be acknowledged by the first on-tree router on the path to the Core.

[10] The remainder of the document reveals that MMP follows the same logic of *macro/micro* mobility split as applied in some other mobility protocols (see section 2.3.3). Historically, MMP was designed with the scope of multicast as its design target before the actual emergence of the *micro/macro* mobility protocols and the associated concepts. However, the layout out of MMP makes it a typical example of such schemes.

In MMP, Mobile IP is used for *macro* mobility up to the FA, which is located at the ingress point of the foreign network domain. This entity is referred to as the Gateway and it is where the conversion between the *macro* and *micro* mobility takes place. From there onward (downlink) *micro* mobility is handled by CBT mechanisms up to BSs. In Figure 3.5 a network domain is shown with a hierarchical topology where the Gateway is also the Core for the multicast part of MMP and acts as a conversion point between Mobile IP and CBT. MHs acquire multicast *care-of-addresses* by contacting the BSs, which then use CBT to create routing trees up to the Gateways (Core). The Gateway then uses Mobile IP to contact the HAs hence deploying the multicasting part of MMP transparently to the rest of the Internet and appearing as Mobile IP. Multicast trees are built using (hop-by-hop) CBT Join Request and CBT Join Ack messages, (assuming network routers are multicast capable) hence creating <*, multicast *care-of-address* group> router states in routers between MHs and the Gateway. Handover procedures consist of transmitting new CBT Join Requests and updating the routing tree to contain the new path from the Gateway to the BS.

The next section presents a detailed explanation of the protocol mechanisms of MMP.

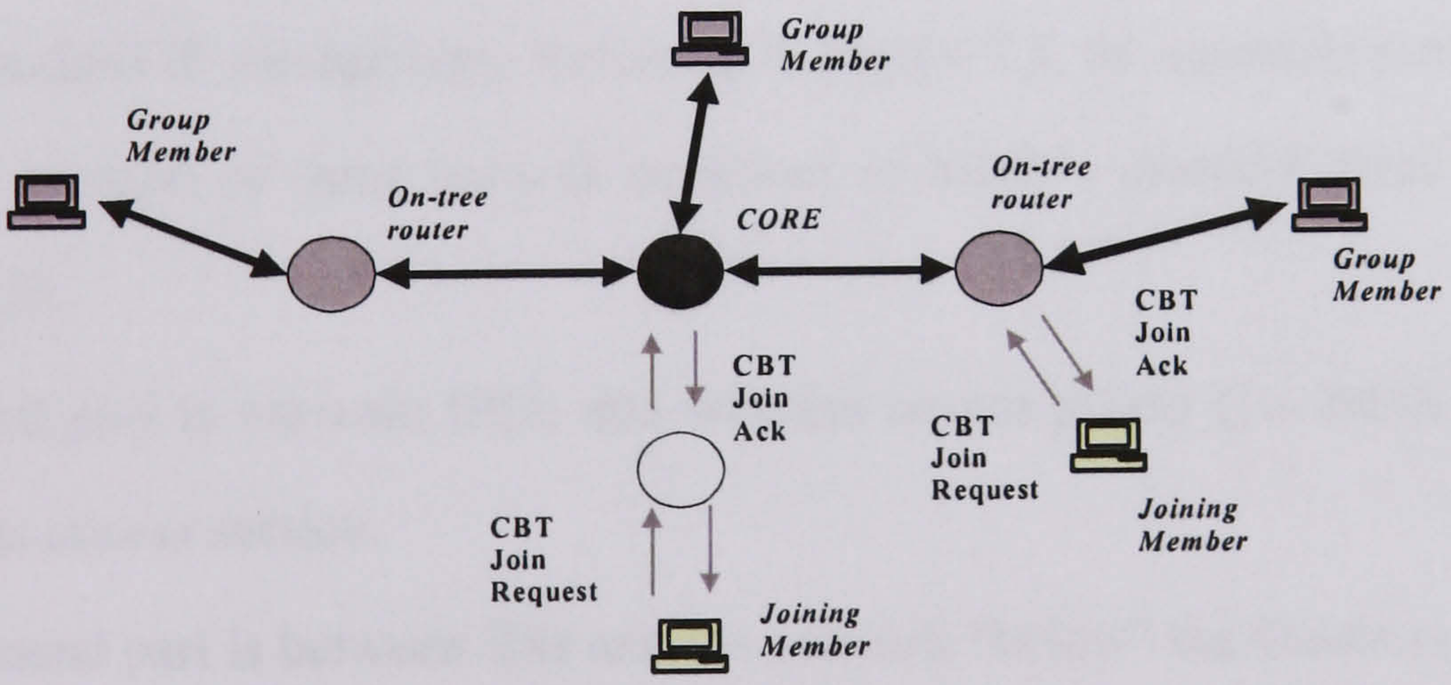


Figure 3.4. Pure IP multicast CBT scenario with established trees and joining routers

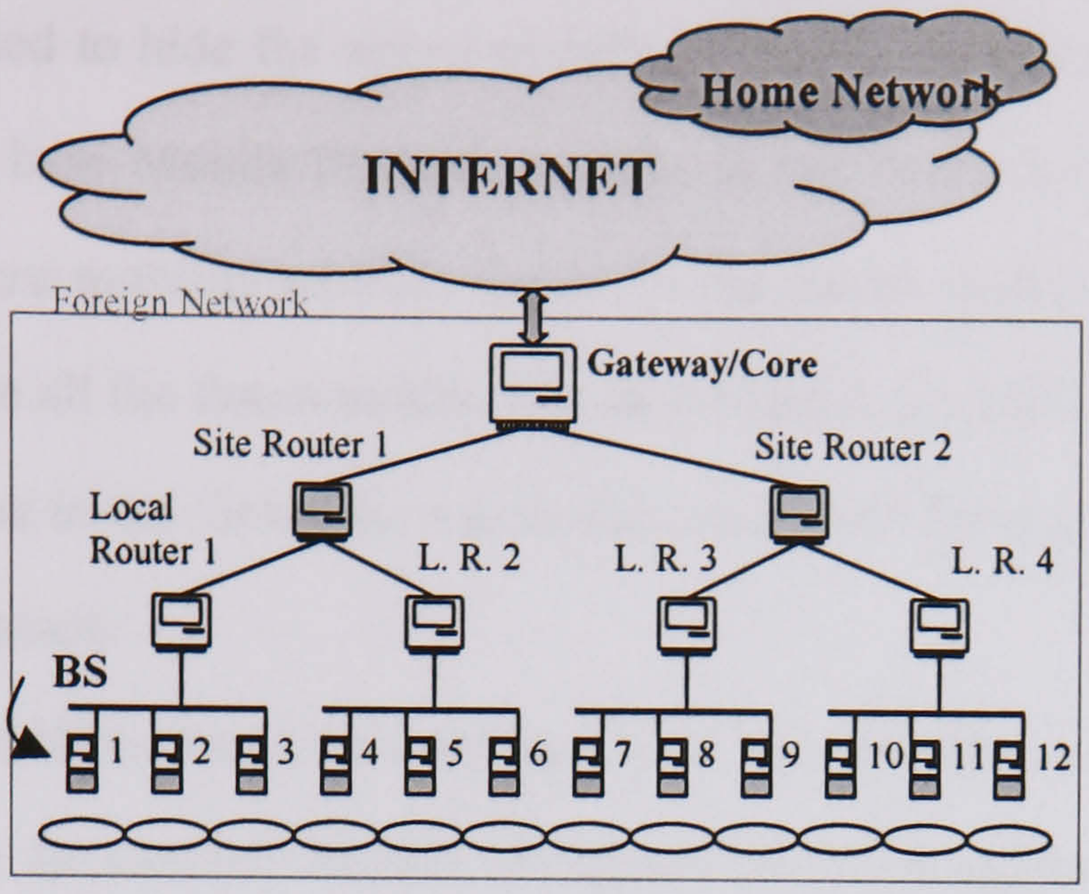


Figure 3.5. An example setup of MMP

3.4 MMP Protocol Setup

3.4.1 Location Management and Routing

It was indicated in the previous section that MMP uses Mobile IP for *macro* and CBT for multicast-enabled *micro* mobility using a network setup conceptually similar to the one shown in Figure 3.5. Identically to most other IP-mobility schemes, MMP only deals with downlink communication, and assumes that the reverse direction is possible through standard IP mechanisms. Referring to Figure 3.5, the essential protocol layout practically consists of three parts (a depiction of MMP's protocol steps is given in Appendix 3):

- The first part is between MHs and wireless access points (i.e. BSs). This is the *wireless access* section.
- The second part is between BSs and the network "below" the Gateway. This is the *micro* mobility section.
- The third part is between the Gateway and the home network of a MH. This is the *macro* mobility section.

MMP is intended to hide the *micro* mobility from the rest of the Internet and appear transparent, as base Mobile IP, both to MHs in the *wireless access* and to HAs and CHs in the *macro* mobility section. Hence, in the *macro* mobility section, Mobile IP is required to have all the functionality: HA in the home network and a FA, which in this case needs to be in the Gateway, where the conversion between the *macro* and *micro* mobility takes place.

Since *micro* mobility should not be visible in the *wireless access* section, MHs are expected to use the standard Mobile IP Agent Discovery mechanisms to obtain a *care-of-address* and establish a link-layer awareness. This requires BSs to appear as FAs to MHs, although as far as MMP is concerned they are required only to have all functionalities of standard IP routers. BSs include Agent Advertisements in their periodically transmitted wireless beacons or alternatively respond to Agent Solicitation messages sent by MHs as indicated by Mobile IP. MMP uses multicast addresses as *care-of-addresses* and assumes that this does not affect Mobile IP mechanisms in MHs, which are expected to send Registration Requests to serving BSs after the Agent Advertisement procedure is completed containing the obtained *care-of-address*.

The *micro* mobility section of MMP is initiated when a BS receives a Registration Request from a newly connected MH. The BS should forward the Request to the Gateway where the conversion to *macro* mobility takes place. The Gateway replaces the *care-of-address* with its own address and relays the Registration Request to the HA. Thus, it appears to the rest of the Internet as a FA of a MH and receives all packets sent to it. Apart from forwarding the Requests, the BS has to start the process of creating the routing tree in the *micro* mobility section, so that packets arriving at the Gateway can be routed to the MHs. Upon completion of the initial step of MH's registration to BS, the BS should transmit a CBT Join Request up to the Gateway, which should then acknowledge it with a CBT Join Ack. The CBT Join Ack traverses the reverse path of the Join Request all the way "downstream" to the BS. CBT Join

Requests create ***transient*** router states up to the acknowledging point, in this initial case the actual Gateway. CBT Join Ack messages, which traverse the reverse path of Join Request messages, create ***permanent*** router states <*, *multicast care-of-address*> in routers between the Gateway and the BS. As specified by CBT, Join Requests are addressed and sent to the Core of a multicast group. In the MMP setup, the Core is collocated with the Gateway, which is the pivotal point in the system being equally spaced between BSs and which acts as the conversion point between *micro* and *macro* mobility sections. Hence, the multicast routing architecture formed for MHs resembles a partial CBT setup where the Core is the Gateway and any BS are acting as leaf (local) multicast routers. It is important to note that MMP uses Registration Requests to trigger multicast tree forming/joining and not the standard IGMP, which is used for communication between multicast joining hosts and local multicast routers, most commonly through the link-layer connections.

Packets, originally addressed to the MH's home address, are intercepted by the HA, which tunnels them to the Gateway (seen as a Foreign Agent). The Gateway decapsulates the packets and encapsulates them again in multicast packets which are sent along the CBT tree formed down to the BS which finally decapsulates them and delivers the original packets to the MH.

The use of a multicast *care-of-address* per single MH introduces a problem during their assignments since it is crucial that every BS always broadcasts a unique multicast address, which is not currently being broadcast elsewhere in the network nor currently used by an active MH. Broadcasting an already active multicast *care-of-address* would create ambiguity in the creation of overlapping multicast router entries for different MHs. One possible implementation solutions could be to have a centralised pool of multicast *care-of-addresses* that can be queried each time a BS assigns a multicast *care-of-address*, and requires a new one for future MHs.

3.4.2 Handover

Handover is initiated by MHs and is in accordance with one of the movement detection algorithms specified by Mobile IP (see section 2.3.1). MHs can, by examining received beacons (eg. through Prefix-Lengths Extensions), determine the presence of another BS and send a new Registration Request to it. This registration step is a deviation from Mobile IP procedure, since the request message contains the old multicast *care-of-address*, whereas in Mobile IP new Registration Requests always contain new *care-of-addresses* due to the change of subnets[11] (since BS are wireless IP routers, every cell is a subnet). The BSs are not FAs but relaying routers as far as the Registration Request message is concern, apart from the initial login when they "hand out" the multicast *care-of-address*. Thus, upon reception of a request message BSs should simply forward the message to the Gateway (actually the message is forwarded to the HA) regardless of whether the MH is setting up the initial connection or performing a handover.

The Registration Request includes the multicast c*are-of-address* previously obtained (provided the MH is still in the same Gateway-scoped foreign network), which is propagated to the Gateways and triggers the multicast tree joining as indicated in the previous section. The Gateway can then compare the new Registration Request with the previous one and, only if different, forward it to the HA to achieve consistency of the *macro* mobility connection. In this way the benefits of *micro* mobility are preserved and the Internet is not overloaded with unnecessary Registration Requests and Replies since it is highly likely that MHs will transmit the same Requests during handovers.

In the *micro* mobility section, the multicast tree needs to represent the routing path from the Gateway (Core) to the new BS. CBT provides a simple solution during

[11] A similar scenario is present in HAWAII protocol where MHs have CCoA for the whole of the connection lifetime in a network domain. During handovers they preserve their CCoA. An extra flag in the registration message may be used to indicate such procedure.

handovers by forming a new routing branch only up to the "**cross-over**" router (see Figure 3.6), which is the nearest router on the path to the Core (Gateway) and the previous routing tree. The previous routing tree included the path from the Gateway to the old BS. The tree forming mechanism is the same as explained in the previous chapter and includes sending a Join Request message, which is acknowledged by the "cross-over" router with a Join Ack message.

Referring to the network in Figure 3.5 and Figure 3.6, there are three types of handover distinguished by the ***handover distance***, defined as the number of hops a CBT Join Request traverses before the "cross-over" router is reached. The *handover distances* in the network setup shown in the figure can take values between one and three. As an example, handovers between BS 2 and BS 3, BS 3 and BS 4 (see Figure 3.6) and BS 6 and BS 7 have distances of one, two and three respectively. The biggest delay will occur for a *handover distance* of three, which is the handover between BS 6 and BS 7 since the Join Requests travels all the way up to the Gateway (Core) due to the previously formed tree including the Gateway, Site router 1, LAN router 2 and BS 6. The essential characteristic of the handover mechanism is the small route update delay, which even for maximum *handover distance*, performs better than the Mobile IP registrations.

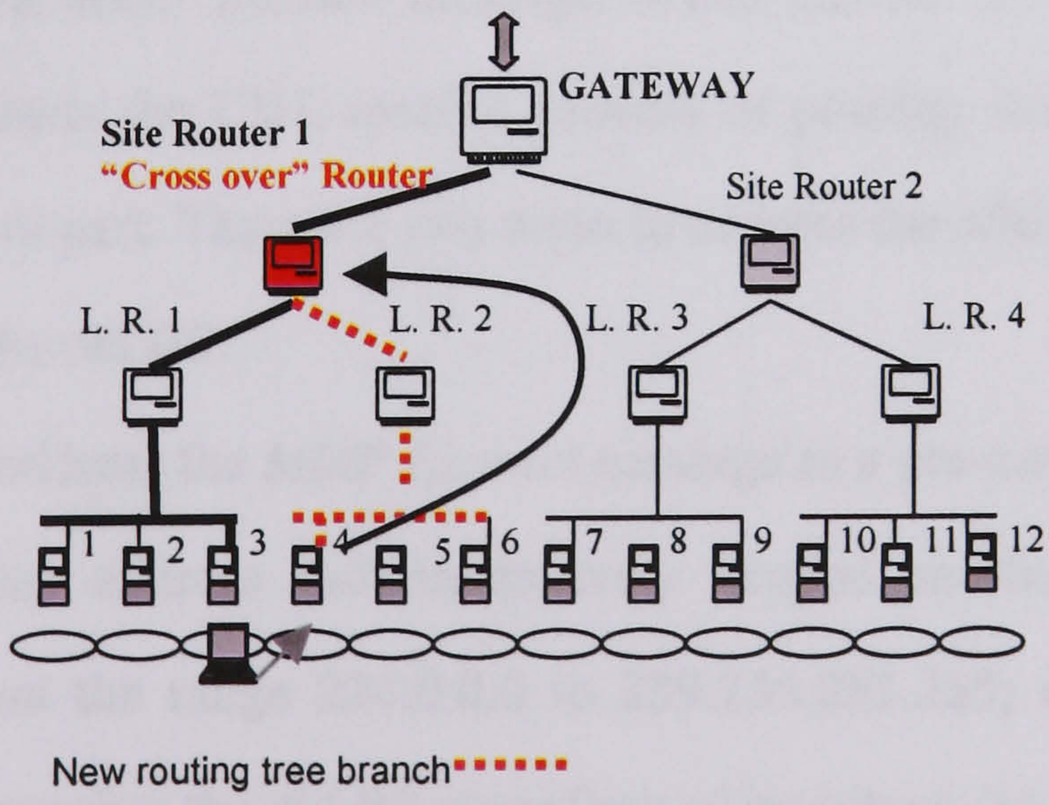


Figure 3.6. An example handover between BS 3 and BS 4 where *handover distance* = 2. "Cross Over" router is Site Router 1.

Due to the nature of CBT, which is primarily a multipoint (i.e. multicast) communication protocol, old branches of the multicast tree used for routing in MMP are not cut-off after a handover. The "cross-over" router, even after the reception of the new Join Request from the new BS, still keeps the old entries pointing towards the old BS and relays packets to both the new and the old BS. This naturally wastes bandwidth in the old routing branch and in the old wireless cell of the old BS. Unlike Mobile IP where entries in the old BSs (FAs) are not active as soon as the HA gets updated, in the case of MMP the old BS is still a part of the routing tree and remains so until it is explicitly pruned or after a timeout. The timeouts of old multicast trees are by default in the order of minutes, which is unacceptable considering the possible packet "wasted" in non-active routes. Since IGMP is not used in MMP (although there is an IP multicast portion of the protocol and additionally IGMP can provide a *Leave* message to "cut-off" old tree entries) an alternative solution would be to propose a new control message which would cause the old BSs to delete entries and "cut-off" the old tree branch up to the "cross-over" router. The solution adopted is to have the new BSs transmit a new control message, called *MMP Instruct*, to old BSs through the wired link. This is consistent with the design goal of making MMP-specific mechanisms transparent to MHs since they are not required to participate in the process of sending the *MMP Instruct* message. When the old BS receives the *MMP Instruct* message it starts the CBT-specific process of pruning the unused branch of the tree from the active part. There are two ways to address the *MMP Instruct* message from the new BS to the old BS:

- The new BS can address the *MMP Instruct* message to a pre-configured "*all-Base-Stations*" multicast address (administratively scoped multicast addresses: an available one from the range 224.0.0.0 to 239.255.255.255; see section 1.4) to make sure that it reaches the old BS regardless of its subnet (an alternative to "*all-Base-Station*" address would be "*All-Neighbouring-Base-Stations*" requiring a specific configuration for each BS…).

- Instead of overloading the network with an administratively scoped multicast messages, which would inevitably be received by all or a portion of BSs, a more resource saving solution would be to address the *MMP Instruct* message directly to the old BS. Information about the address of the old BS should come from MHs. Various solutions are possible, both in the link and the network layer, but the one that is most consistent with the principle of transparency of *micro* mobility is to use the Previous-Foreign Agent Notification option in order to indicate the address of the old BS as indicated by Mobile IP Route Optimisation [27] (addresses of BSs can be deduced from Agent Advertisements). This includes attaching a Previous-Foreign Agent Notification extension to the Registration Request thus instructing the new BS to send the *MMP Instruct* message to the old BS. The procedure in MMP is different to the Mobile IP Route Optimisation where the BSs would also exchange binding *messages* (see section 2.3.1). This is intentionally avoided to reduce the unnecessary complexity of MMP's operation, which relies on other mechanisms in the *micro* mobility section. Additionally, as specified by Route Optimisation, Mobility Agent Advertisement messages transmitted by BSs are assumed to include the additional "S" flag to indicate that they support this feature. This addressing solution is adopted for MMP.

Informing the old BS that the associated routing tree branch is stale reduces the *leave latency*, defined as the time between the invocation of the transmission of a *MMP Instruct* message to the time taken to completely remove the old routing branch and stop forwarding packets to the old BSs (see Figure 3.7).

The actual time of transmission of the MMP *Instruct* message can be varied in order to adjust the style of handover. In the current MMP design an *MMP Instruct* message is transmitted as soon as the old BS receives the Registration Request from the entering MH, unless the Request had the "S" flag set, in which case the MH is requesting *advance registrations* (see next section). Upon receiving the MMP *Instruct* message the old BS removes the non-used branch of the tree by using a CBT Quit

Notification message, which removes the old branch of the tree up to the "cross over" router. The *MMP Instruct* message has an ICMP message format and contains an adequate multicast *care-of-address*, which will cause the old BS to delete the old entry. To ensure that MMP *Instruct* reaches the old BS it should be retransmitted few times.

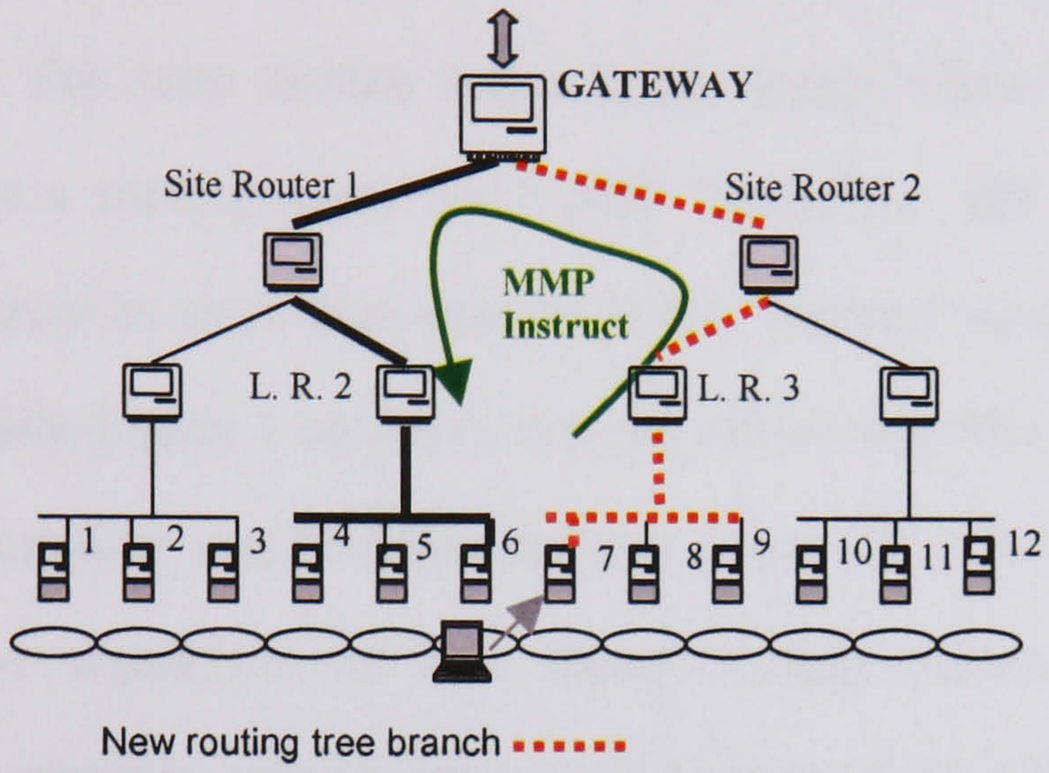


Figure 3.7. Path of *MMP Instruct* message after a handover between BS 6 and BS 7 when *handover distance* = 3. "Cross over" router is the Gateway.

3.4.3 Other Protocol Mechanisms

3.4.3.1 Soft State

This applies to the ability of routers to maintain an awareness of the relevant, usually neighbouring routers and adapt to any changes that may arise due to a router or link failure. The "soft state" problem is mainly applicable to the *micro* mobility section since the routing tree is formed there and it is where it needs to be "maintained". Also, the remaining parts of MMP are controlled by the standard IP mechanism.

CBT is deployed in the *micro* mobility section and has its own "soft state" mechanism [18], which fully complies with MMP operational requirements (i.e. it can be used natively as specified by CBT without affecting any of the operations of MMP). The basic operation consists of a router transmitting a CBT "keepalive" message called Echo Request to its first uplink (towards the Core) neighbour, which then

acknowledges it with a new message called Echo Reply. This process is independent for every router (the Core/Gateway only replies to Echo Requests but does not transmit them) in the routing tree and needs to be performed periodically after a specific interval called the Echo Interval specified in the protocol arrangements. The most important characteristic of MMP's "soft state" mechanism for *micro* mobility is that it is not influenced by the number of MHs simultaneously using the same **section** of the network. The term section refers to the same "chain" of routers where each router maintains a routing entry for a MH "under" a BS at the "bottom" of the "chain". There may be more than one BS at the "bottom" of the "chain". Referring to Figure 3.5, the Site Router 1 can have routing entries for MHs in BS 1 to BS 6. Again, there may be more than one MH attached to a single BS. CBT executes the same "soft state" procedure regardless of how many routing entries a router is currently supporting. For example, routing trees could be formed for a large number of MHs per cell but the transmission of Echo messages will not be a multiple of the number of MHs sharing the routing tree section. In fact, they would be the same as if there was only one MH with the routing entries. The same applies for overlapping trees up to the Gateway, which come from different cells where the overlapping section of the routing tree behaves as a routing tree for a single host. The first routing tree established determines the timing of the transmission of Echo messages. **The original "soft state" mechanisms of CBT is intended to verify the status of upstream routers (toward the Core) for the whole multicast architecture and not a single group entry**. Thus the CBT "soft state" mechanism is designed in an aggregated manner for all routing entries. In comparison with other mobility protocols, this feature reduces the protocol overhead as the number of MHs in a network increases. Usually, in most mobility protocols the protocol overhead of the "soft state" increases proportionally to the number of MHs attached (see section 4.3 for simulations of MMP's protocol overhead). In the *wireless access* section, BSs interpret periodic transmissions of Mobile IP Registration Requests by MHs as an indication to maintain

the upstream tree. The possible tearing down of a tree from a BS upstream towards the Core or up to an active section of the tree branch, is performed by the transmission of CBT Quit Notification messages, caused by a timeout or the reception of an *MMP Instruct* message at the BS.

3.4.3.2 Support for Idle Hosts/Paging

Support for idle hosts (i.e. paging[12]) should achieve two goals: reduce the protocol overhead (signalling and memory requirements) in the network and minimise the power consumption for idle hosts by reducing the frequency of protocol refreshments/updates. MMP is not concerned with reducing power consumption for idle MHs since they do not participate in MMP-specific location management and the "soft state" mechanisms (see the previous section). MHs use Mobile IP to maintain established connections, which themselves do not impose large overheads.

The signalling and memory requirements in the access network for MMP are significantly lower than in other *micro* mobility examples such as Cellular IP and HAWAII due to the aggregation of the "soft state". As indicated in the previous section, Echo messages are exchanged between routers in a hop-by-hop manner and are independent of the number of MHs using the routing tree. Support for idle hosts in MMP should take the simple form of reducing the frequency of "soft state" messages for the routing trees. Since MMP's "soft state" mechanisms are common to all users sharing a particular set of routing entries, the frequency of "keepalive" (Echo) messages does not have to be adapted to a particular MH that may be idle at a given instant, but to all MHs using a particular section of the tree. Adjusting the frequency of the "keepalive" messages is done through the adaptation of the Echo Interval by

---

[12] The term paging comes from the terminology and functions existent in the design of cellular systems. However, it is embraced in the design of IP mobility management because most of the schemes deploying the idle mode support actually use a method of "paging" idle MHs and hence invoke a transition to an active state. While the paging in MMP does not extend to actually include a "paging" step as in some IP proposals, the primary goal of idle mode support is the same and the general terminology can be applied. More on this can be found in Section 5.2.2.4.

managing two possible values: *Active* and *Idle* value (the latter should be a large multiple of the former). Routers should use the *Active* value as a default value and switch to the *Idle* value after *x* (*x* specifies the exact number of times the timer can expire) expirations of the *Active*-valued timer without receiving an indication that there is at least one active MH using that particular routing branch. There are two possible indications of an active user in MMP: downlink and uplink packets, addressed to (*care-of-address*) or sent by a MH respectively. Alternatively, MH can transmit periodic Registration Requests in the *wireless access* section as specified by Mobile IP. These transmissions of the Registration Requests are typically infrequent and are not used as an indication of the state of MHs. Besides having a low frequency, the periodic broadcasts by MH are not controlled by a MMP-specific mechanism (since MMP is transparent in the *wireless access* section) and are therefore not included in the idle mode decision making procedures.

Routers can monitor downlink packets sent to MHs since they route them anyway according to the multicast *care-of-addresses* and established routing trees. However, uplink packets are routed using the standard IP routing mechanisms and MMP is not involved. The Gateway (Core) keeps the same timers as the routers and checks for any uplink traffic sent by MHs in the system and, in the same way, routers check the downlink traffic. The requirement is that the uplink packets pass through the Gateway (this is possible in the topological setup of MMP since the Gateway is the anchoring point in the system and all traffic flows through it anyway). In this way the Gateway can deduce whether a MH, and hence the appropriate routing tree, is still active and suppress the invocation of the idle state, even if it is not receiving any downlink traffic. Upon the expiration of the timer, the Gateway should create an MMP *Instruct* message, with no specific multicast *care-of-address*, and send it via the particular branch of the tree "down" to BSs. The tree branch can be deduced from the source address of the uplink packets (home address of the MH), which can then be matched to the particular multicast *care-of-address* and the originating BS, which was also

included in the initial Join Request. For the *advance registration* (see the next section) option where MMP *Instruct* has to reach more than one serving BS the Gateway will have to store all appropriate BSs. The decision to store multiple BSs can be made after receiving a Registration Request with the appropriate flag set (in particular, the "S" flag according to Mobile IP specifications) to include simultaneous entries. Finally, routers interpret this empty MMP *Instruct* message as an active-host indication and recipient BSs destroy it.

The idle mode support in MMP is not critical since the protocol essentially does not impose large overheads due to the aggregation of "soft state". Thus, the "soft state" feature is optional and should not be deployed if the ease of deployment of MMP is to be maintained since additional features would be required in the CBT protocol inside the *micro* mobility section.

3.4.3.3 Advance Registration

Advance registration refers to the ability of the routing protocol to establish active routes to more than one point-of-attachment of a MH. This was one of the primary reasons for the consideration of multicast in mobility protocols because of the natural ability of multicast protocols (as multipoint support protocols) to install more routing branches to different group members (or in mobility terms: different points-of-attachment of single MHs). This means that the routing tree leads to more than one BS and branches off from the "cross over" router. The main benefit of this feature is that a MH can have a seamless handover because the packets destined to it will already be available in the new cell. To perform *advance registration*, MHs should transmit Registration Requests with "S" flag set to indicate simultaneous bindings, as specified in Mobile IP. In this situation the new BS is not required to transmit the MMP *Instruct* messages to the old BS (since the "S" bit is set) because the old branch of the tree is still being used. MHs can cancel the *advance registration* option by retransmitting a Registration Request without the "S" flag set hence causing the BS to

transmit an MMP *Instruct* message to cut-off the old branch. The advance registration feature can be further analysed considering the detailed handover specifications. This is contained in Chapter 6.

3.4.3.4 Support for Mobile Sources

The whole of the explanation of MMP's protocol mechanisms in the previous sections of this chapter relates to the delivery of packets to MHs, from HAs and CHs located outside the *micro* mobility section. As indicated in section 2.2, downlink communication is the essential concern of all mobility protocols. For CHs located inside the *micro* mobility section of the recipient MH, which are sending packets to the MH, it is highly desirable to avoid unnecessary delays which would be caused by sub-optimal routing to the recipient's home network and back. In other words, for this particular setup, the CH and the MH may even be located in the same cell but the packets sent to the MH would travel all the way up the MH's HA and then back to the same cell before they reach the MH. A straightforward solution to this problem is to have the Gateway check all outgoing packets and compare their destination address with the home addresses of MHs currently supported by MMP in the *micro* mobility section (a routing entry always contains the *care-of-address* and the home address of a MH). If a match is found, meaning the recipient is actually in the network, the Gateway should reverse the packets' movement back inside the network (i.e. downlink), as if they had arrived from the HA of the CH. Packets are then normally forwarded to the destination (i.e. MH) inside the MMP-scoped network using the MMP routing tree entries. Gateways are easily instructed to perform this by adding additional functionality in the MMP-specific filtering modules.

3.4.3.5 Implementation scenarios

Since MMP mostly relies on existing Mobile IP and CBT; any new MMP-specific features are only needed at the transition points between the two sub-protocols. These

points are between the *wireless access* and the *micro* mobility sections and between the *micro* mobility and *macro* mobility sections, that is, in BSs and the Gateway respectively. The Gateway must be enhanced with MMP-specific filtering and control modules to perform tasks such as: decapsulation and encapsulation between *micro* and *macro* mobility sections, processing of Registration Request and Replies according to MMP design and other relevant tasks explained in this chapter. Practically, the FA-representing module in the Gateway needs to interact with the CBT-part of the routing kernel and accordingly transfer packets between the two parts of the protocol (see Figure 3.8). BSs should interpret Registration Requests as IGMP messages in native multicast and in response trigger CBT tree joining, again, by providing the "gluing" functionality between the multicast and Mobile IP-representing mechanisms. BSs should also generate or accordingly receive, *MMP Instruct* messages and invoke the appropriate CBT processes.

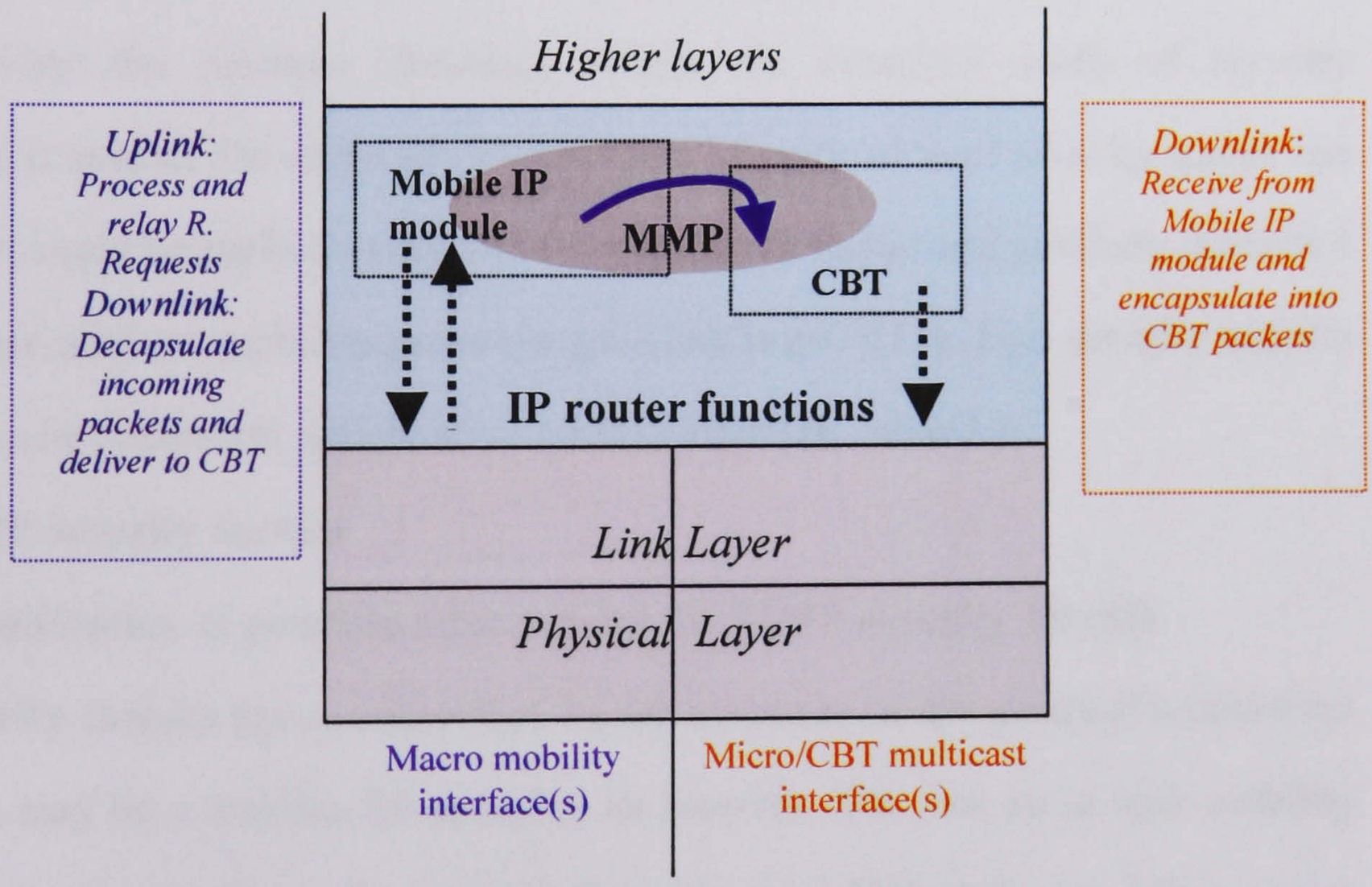


Figure 3.8. Transitional features of MMP in a Gateway for delivery of packets to MHs and uplink/downlink delivery of control messages

One of the design goals of MMP is to retain the CBT multicast portion of the protocol in the *micro* mobility section without any additional MMP-specific mechanism apart from the ones needed in the transition point explained above (Note: as indicated the "soft state" feature is optional).

The primary reason for retaining the multicast portion of MMP is to make the eventual protocol implementation flexible because CBT is an existing protocol in the Internet and can easily be implemented. However, MMP can be realised independently, without the multicast, where all messages and their use can be identical to the control messages of CBT and the related protocol mechanics, but without the dependency on multicast and the multicast *care-of-address*.

3.4.3.6 Security

Securing the operations of a mobility protocol is essential for its success and viability as a deployment candidate. This applies to both the wired and wireless segments where mobility the protocol functions. While the complete study of security mechanisms is outside the scope of this research, identification of security issues and methods that could be applied as solutions is considered useful and similarly described in some other relevant mobility protocols [28][35][36][47][71]. Two security aspects for MMP can be addressed, which are also summarised in Table 3.1:

a) **MMP security threats**

b) **Identification of possible solutions for the MMP security threats**

MMP security threats are considered to be the instances of the protocol's operation where there may be a concern for securing its features. The first issue with mobility protocol is allowing access to the network for authorised MHs only, i.e. *MH's access to the network*. Hence, MHs are required to produce a proof of their identity before they can be authorised to use the network and run the MMP features. Once this is resolved, *MH generated control messages* over the wireless link (these are Mobile IP Registration Requests as specified in the previous sections of this chapter) trigger

MMP tree establishment in the foreign network. The network must trust these messages to prevent occurrence of malicious attacks when another party generates the same messages and causes erroneous tree establishment. These messages have to be authenticated by the network in order to trigger *creation of MMP micro-mobility routing trees* (i.e. CBT multicast tree creation) for delivery of packets to MHs. In addition to authentication, *replay* protection may be needed in situations when another party copies the whole or part of the content of the message transmitted over the wireless link by MHs and generates another forged message using the content of the original message. Regarding the *exchange of control messages between MH and its HA*, the messages also need to be authenticated and protected against replaying. Another security threat is the *Gateway's processing of control messages*, that is, the Registration Requests and Registration Replies and insertion of its address as the care-of-address conveyed to MH's HA. This requires authentication between the Gateway and MH's HA, in order for HAs to trust the Registration Requests processed by the Gateway. Furthermore, firewalls and associated packet *filtering* in the foreign network should allow packets to be sent to and from MHs, which are authorised to use the network.

Identification of possible solutions for the MMP security threats is significantly related to distribution of security data to all relevant entities of the protocol setup. This data, commonly referred to as the shared secret [35] (or shared secret security key or in some scenarios security association containing additional data alongside the key such as the required security algorithm [71]), would then be used for overcoming a large portion of the MMP security threats. Initially, authorisation of *MH's access to the network* can proceed as a step preceding MMP's mobility functions where MHs would supply their credentials in order to be allowed to use the network. One such model is already applied for Mobile IP [71] and recommended for other protocols such as Hierarchical Mobile IP [33], HAWAII [37] and BCMP [47] (this model can also be adopted for MMP since Mobile IP is used for macro mobility and always for

MH generated signalling). The model is based on contacting a local AAA (Authentication, Authorisation and Accounting) authorisation authority, which can then contact the home AAA authorisation authority for the MH (using a suitable signalling protocol) in order to validate the MH's credentials and derive the shared secret. In Mobile IP, the local and home authorisation authorities are assumed to coexist or interoperate with Foreign and Home Agents. In HAWAII the home authorisation authority is also present in HAs while the local authorization authority is assumed to be the Gateway of the foreign network. The same can be assumed for MMP, i.e. the local authorisation authority could coexist or interoperate with the Gateway.

MMP Security Threats	Description	Recommended Solutions
MH's access to the network	Allowing access and use of the network for authorised MHs only	AAA-based solution [71] where local AAA authority contacts MH's AAA authority to check MH's credentials and assist in deriving the shared secret
MH's generated control messages	BSs need to trust the messages generated by MHs to proceed or trigger further actions	Shared secret used to provide authentication for BSs. For availability of the shared secret in new BS few solutions possible (reactive/proactive)
Creation of MMP micro-mobility routing trees	The established routing tree only for authorised MHs and to their correct current point of attachment, i.e. BS	Scoped and generated by foreign network routers which can trust each other or have preinstalled security mechanisms
Replays	To prevent malicious users from copying content of MH's control messages and regenerating erroneous messages	Mobile IP solution [28] using timestamps or random numbers. MH messages could be further encrypted over the wireless link to conceal the content
Exchange of control messages between MH and HA	For macro-mobility and to certify the current foreign network and MH's care-of-address	Mobile IP [28] solutions for authenticating the messages. HA has a preinstalled security association with MH
Gateway's processing of control messages	MH's HA needs to trust any messages processed by Gateway and data supplied in them	Requires HA to trust the Gateway. Can be provided in the initial AAA step for exchanging necessary secrets if needed
Filtering	Only authenticated MHs can use the network and send/receive packets	Reverse Tunnelling used in Mobile IP or updates of the firewalls for authorised MHs

Table 3.1. Summary of MMP Security

Securing the *creation of MMP micro-mobility routing trees* is significantly relaxed due to surrogate generation of CBT control messages in the micro mobility domain. BSs trigger CBT messages transmission after receiving Mobile IP control messages (i.e. Registration Requests) sent by MHs over the wireless link. Hence, none of the CBT control messages (including the *MMP Instruct* message) are generated or processed by MHs but entirely by the network, MHs only trigger them. The security threats regarding the multicast forwarding management are relaxed due to the fact that the network would naturally trust or have preinstalled security mechanisms for authenticating the control messages generated by its routers (the same issue is recognised in Cellular IP [36] and HAWAII [37] regarding the exchange of messages inside the wired part of the foreign network). In addition, micro mobility mechanisms are entirely scoped by the wired part for the foreign network. The issue with applying the shared secret to authenticate any control message in MMP, i.e. *MH generated control messages*, is related to transmission of Mobile IP Registration Requests by MHs, for example, for allowing execution of handovers as specified in section 3.4.2. As with other micro mobility protocols the issue is concerned with availability of the shared secret at new BSs when MHs perform handovers and send Registration Requests to new BSs, where the shared secret is included in the authentication data (if not entirely encrypted). A straightforward solution is to have the local authorisation authority (e.g. the Gateway) or an instrument of it in the network supply the shared secret to each BS upon the handover to it. This may cause additional handover delays. Some other mobility protocols have adopted faster ways of delivering shared secrets to new BSs. One solution is to apply the model adopted for Cellular IP [35] where the network (Cellular IP places this functionality in the Gateway) initially uses a specific method for calculation of the shared secret based on some of the MH's known credentials and a security key known to all routers and BSs in the network. The shared secret is then used for authentication and can be decrypted by all BSs automatically. Another method for immediate availability of the shared secret in BSs is to distribute

the shared secret to all BSs in the network so it is already available when MH performs a handover. A variant to this could be implemented by sending the shared secret only to surrounding BSs in order to reduce the overhead. Some other solutions, which go beyond the MMP specifications in this chapter, can include transfer of the shared secret during the handover steps as proposed in Chapter 6. In addition, new BS can receive the shared secret from the old BS during the handover in an acknowledgment message to the *MMP Instruct* message.

Regarding the processing of Registration Requests and Registration Replies by the Gateway, the Gateway need to share the secret with MHs and also with their HAs. Sharing the secret with MHs is provided already during the authorisation and authentication of MHs during their login to the network as already used in the above mentioned procedures. Negotiating a shared secret with the HA can also be done during the MHs authorisation when the foreign network (i.e. the Gateway) checks MH's credentials with its home network (home authentication authority in HA) using the AAA model for Mobile IP [71] explained before. In this way, *exchange of control messages between MH and HA* is covered by Mobile IP mechanisms so the *Gateway's processing of control messages* can be secured. Hence, the Registration Requests processed by the Gateway and relayed to HA, can be trusted and used for packet forwarding since the Gateway can authenticate itself to the HA. The requirement for MMP authentication between MHs, the Gateway and HAs already fits into the models used for generation of shared secrets explained above [71]. Mobile IP [28] applies the obtained shared secret by having three types of authentication pairs for exchanged messages: mobile-home, mobile-foreign and foreign-home authentications. In MMP, the first part is applicable to the authentication between MHs and its HA, the second pair is applicable to the authentication between MH and the Gateway and the third pair is applicable to the authentication between the Gateway and HA. The Gateway is functioning as the agent in the foreign network responsible for handling authentication of MHs. This already goes along the functionality needed in the Gateway specified in

MMP which appears as a Foreign Agent of MHs to the rest of the Internet (see previous section).

MHs needs to generate Registration Requests with correct *replay* protection solutions to certify their validity. Timestamps or newly generated random number (a nonce) can be used as already specified by Mobile IP [28] and included in Registration Requests and Registration Replies. Regarding the wireless link, MHs can use the shared secret not only to provide authentication of the messages (by providing authentication extension) but for encrypting the whole message sent over the wireless medium which would make its content hidden to any malicious attacks (however, replay protection is still needed as explained above. In addition, the network can renegotiate the shared secret periodically to further increase the security).

An additional thing could be important for making MMP operational in secure environment is the firewalls and *filtering* of packet that pass through the network that may be present in foreign networks where network may discard uplink packets which have source addresses belonging to a different network (i.e. MH's home network). This can be avoided if MH's uses reverse tunnelling to its HA (encapsulation back to HA before delivery to CH) as already suggested in Mobile IP using the multicast care-of-address or another address from the foreign network; or foreign network can update the firewall agents with MH's addresses for duration of its connectivity in the foreign network.

3.4.3.7 Summary of MMP Messages, Timers and Features

The following Table 3.2 contains a summary of all control messages used in MMP, their use in MMP, protocol from which they are adopted and associated timers.

Message Name	*Origin Protocol*	*Default Use*	*MMP Use*	*Other Properties*
Join Request	CBT	Creates multicast tree. Sent towards the Core. Creates "*transient joint state*".	Creates MMP routing tree to BSs. Single member of the group – MH. Functionally similar to default use. Sent upstream	Triggered in BSs upon reception of Registration Requests sent

			towards the Core/Gateway.	by MHs. Special retransmission in bus links.
Join Ack	CBT	Acknowledges multicast tree creation. Sent back to the joining member. Creates "*permanent join state*".	Functionally similar to default use. Acknowledged by Core or "cross over" router.	
Echo Request	CBT	Periodically sent "keepalive" (soft state) message. Sent to the next hop upstream for the entire set of entries – aggregation of signalling.	"Soft state" maintenance in the micro-mobility domain. Functionally similar to default use (Aggregated – for all MHs using the hop).	Default transmission period ECHO_INTERVAL = 60 seconds.
Echo Reply	CBT	Acknowledges Echo Requests. Sent as a response next hop downstream. Includes the response for the entire set of entries – aggregation of signalling.	"Soft state" maintenance in the micro-mobility domain. Functionally similar to default use (Aggregated – for all MHs using the hop).	
Quit Notification	CBT	"Prunes" multicast tree downstream-to-upstream, i.e. towards the Core. Timeout or IGMP triggered.	"Prunes" the old tree branch from the old BS. Triggered by *MMP Instruct* after handovers.	Special handing for bus links. Transmitted (default) MAX_RTX = 3 times.
Flush Tree	CBT	For routers tearing down downstream multicast tree. Sent downstream.	Functionally similar to default use. Not affected by MMP functionality.	Not used in Chapter 4 simulations. No instance of router or ink failures.
Registration Request	Mobile IP	Registering and/or refreshing CoA entry in HA (MH-to-HA).	Initially sent by MHs to register multicast CoA. Intercepted by Core/Gateway then relayed to HA including Gateway's address.	
Registration Reply	Mobile IP	Acknowledges Registration Requests (HA-to-MH).	Sent by HA to Gateway. Relayed to MH to confirm registration when necessary.	
Agent Advertisement	Mobile IP	Period network layer beacons sent over the wireless link by BS/FA.	Functionally similar to default use. Multicast CoA.	Used for movement detection in MMP
Agent Solicitation	Mobile IP	Solicitation of Agent Advertisements. Sent by MHs.	Functionally similar to default use.	Used for movement detection in MMP
MMP Instruct	MMP specific	…	Triggers tearing down of the tree from old BS. Sent by new BS. Also for	

			Supporting Idle MHs sent by Gateway downstream.	

Table 3.2. Control messages and their use in MMP

3.5 Discussion on MMP protocol mechanisms

The multicast scoping issues and the choice of multicast routing protocol were discussed in sections 3.2.3 and 3.3. MMP was accordingly designed using the principles of *Multicast terminated Mobile IP* solutions with the *sparse* mode multicast as a routing solution in the *micro* mobility section. However, MMP contains various protocol design choices, which are influenced by other issues of multicast and mobility.

In order to maintain the flexibility of MMP's deployment, the protocol does not put any functional requirements on MHs and expects them to execute the standard features of Mobile IP in the *wireless access* section. The functionalities required in MHs are extended to contain the Mobile IP Route Optimisation feature for the handover model adopted. Thus compared to the existing multicast-based mobility solutions, MHs are not required to use IGMP and are not required to participate in any multicast procedures.

The transparency of multicast in the *wireless access* section of MMP coincides with desired TCP support for MH as end hosts. Since multicast is used in the tunnelled form, that is, inside the *micro* mobility section where BSs perform decapsulation before the final delivery to MHs, there is no need for alterations of TCP codes. These alterations would be necessary if one of the session addresses was a multicast address. In fact as far as the support for transport layer protocols is concerned, MHs are treated (as in the Mobile IP case) as hosts residing in their home network where the network layer mobility functions are completely hidden mainly due to the functionalities of the Mobility Agent and the manner in which CBT is deployed. Additionally the use of

multicast is transparent as far as the network management functions are concerned (i.e. all problems present in MSM-IP) because the use of multicast as the source address of the packets is nonexistent. This means that a temporary unicast IP address is not required as a network management remedy. For example: an ICMP error messages will be sent based on the MH's home address which is the source address of the packet and will be rerouted to the MH from the HA (also address resolution in wireless cells can be performed normally).

3.6 Adaptation of MMP for Internet Protocol version 6

3.6.1 Background

The somewhat unpredicted expansion of the Internet[13] is expected to cause exhaustion of the available IPv4 address space in the foreseeable future. In response to that, the core of the new version 6 of the protocol (IPv6) has already been developed [73]. The main trigger for the work was the requirement for a larger address space. This was achieved by the creation of a new 128-bit IPv6 address [78] and the addition of some new features to it (address "scope" field, anycast addresses…) aimed to achieve greater flexibility and provide autoconfiguration functionalities. Besides the addressing extensions, IPv6 designers have used the opportunity to enhance the IPv6 capabilities with some new features, which compensate for some of the shortcomings of IPv4:

a) Although larger in size (40 bytes) than the IPv4 headers, mainly due to the 16-byte addresses, IPv6 headers contain a simplified set of fields for efficient processing.

b) The IPv6 header has been enhanced by some extension headers, which are to be smoothly integrated into the IPv6 infrastructure. The flexibility of the extension

[13] Already in 1994, 32000 networks were connecting 3.8 million users in more than 90 counties. The expansion seems exponential.

headers and their use enables the inclusion of new features, which may be utilised by the emerging IPv6 protocols.

c) Extended capability for enhanced flow control of packets and in-built security features.

3.6.2 Impact of IPv6 features on mobility mechanisms

Mobile IPv6 is [74], again, the reference mobility protocol in IPv6. It is being designed following the same principles applied in the development of IPv6, that is, Mobile IPv6 capitalises on the possibilities offered by IPv6's new features and experiences from the development of Mobile IPv4 and its shortcomings. Some essential differences in operations of Mobile IPv6, compared to its IPv4 predecessor, are:

a) In-built Route Optimisation to overcome "triangular routing": Mobile IPv6 has combined the registration messages for updating HAs and CHs.

b) There is no need for Foreign Agent functionality in visited networks: MHs are able to configure their *care-of-addresses* based on the available configuration mechanism in IPv6. Two types of method for obtaining *care-of-addresses* can be used: stateless and stateful. Stateless address configuration allows for automatic creation of a *care-of-address*, typically by combining host identifiers and subnet prefixes in a foreign network through the monitoring of Router Advertisement messages (analogous to the Agent Advertisements of IPv4). This feature is available by using the autoconfiguration protocol [75] and Neighbour Discovery procedures [76] in IPv6. Additionally, hosts can obtain a *care-of-address* through a statefull configuration by running a specific configuration protocol such as DHCPv6 [77].

c) Mobile IPv6 benefits from the IPv6 extension headers: The Destination Option header (according to the IPv6 specification, the extension header is used to convey

a particular instruction to a packet's end destination) is used for: containing the home address (Home Address option) of the MH allowing it to use the *care-of-address* as the source address of a packet sent to the CH (hence avoiding ingress filtering and easing the support of other IP protocols) and allowing a possibility of including the control packets (binding messages) in the regular IP packets thus reducing the protocol overhead of Mobile IPv6 signalling (optional feature). Additionally, the Routing header is used for routing packets from CHs to MHs, hence replacing the tunnelling used previously with source routing (HA is still required to perform encapsulation to avoid in-flight modifying of packets).

d) Other features such as: dynamic HA discovery, generic-ARP feature embedded in Neighbour Discovery protocol, utilisation of IPSec security requirements for securing the control message between the routing entities…

Although intended to improve on some of the inefficiencies of Mobile IPv4, Mobile IPv6 is still a protocol mainly suitable for Global mobility due to the registration delays also associated with the previous version. The particular work in developing IPv6 Regional mobility protocols has been focused mainly on adaptations of IPv4 mobility protocols to the new version of IP. Examples include Hierarchical Mobile IPv6 [33] and Cellular IPv6 [79]. Additionally, significant effort is being directed towards the development of fast handover scheme(s), which are proposed as "patch" protocols, to improve the performance of Mobile IP during handovers (see Chapter 6). The adapted IPv6 protocols have been designed largely in a similar way to Mobile IPv6, that is, as an improvement to their IPv4 versions by utilising on some of the functional benefits of IPv6. As an example, Hierarchical Mobile IPv6 uses a relaxed addressing method for MHs and easier transfer of control messages. Cellular IPv6 is also proposes minor changes to the original protocol: local control messages can be inserted in the IPv6 Hop-by-hop extension headers, security feature of IPv6, stateless address configuration and changes in the transitional points between the Cellular IP-specific Regional mobility management and Mobile IPv6 part of the protocol

(Gateway and BSs). The general conclusion from the new IPv6 mobility protocol is that while some of the mobility features are integrated more smoothly than in the IPv4 mobility protocol, the general flow of control messages and their impact on the overall protocol performance has remained the same. In fact, it could be stated that the general logic used in the explanation of IPv4 mobility protocols, can still be applied to the IPv6 mobility protocols available.

Although not detailed in this Chapter (and in the next one where results are presented for the MMP designed for IPv4), MMP has been adapted to IPv6 and an initial version of the protocol has been created and simulated [7]. Performance wise, MMPv6 performs identically to MMPv4 experiencing the same *handover distances* for identical handover scenarios. Apart from the applications of same principles of macro mobility controlled by Mobile IPv6 and micro mobility controlled by MMP-specific internal routing, the protocol relies on only a few assumptions and changes from the IPv4 version:

- CBT is assumed operational in IPv6.
- Since Mobile IPv6 does not deploy FAs, none of the MMP routing entities are required to deploy any of the FA's functionality, or appear as such to any other entity in the protocol. However, the Gateway/Core still needs to perform encapsulation of the downlink packets into multicast packets. This may differ according to where the packets are coming from: CHs use Routing headers, HAs use encapsulation. An additional step may be performed if packets are received from a HA, where the Gateway is the tunnel end-point and it decapsulates the packets before encapsulating them again in multicast packets (as done in MMPv4, Gateways need to replace the *care-of-address* of the MH's binding update message [analogous to Mobile IPv4 Registration Requests] with their own address and appear as MH's to the HA). For the binding updates sent to CH, the Gateways do not change the source packets of the messages so CH stores the

exact address of the MH. A similar method is used in Cellular IPv6 where the Gateway is equipped with additional control and filter modules to enable forwarding of packet with respect to Cellular IPv6 local routing.

- MHs obtain a unicast *care-of-address*. The multicast CBT routing is still performed by using a multicast address as the group identifier, but in IPv6 it is not assigned to the MH since address space exhaustion is not a significant issue. Thus the multicast routing is done in a surrogate way to the Mobile IPv6 signalling. Agent Advertisement/Solicitations are replaced with Router Advertisements/Solicitations of Neighbour Discovery in IPv6.
- BSs interpret the Mobile IPv6 binding updates as triggers for starting the multicast tree creation process. This is done in a surrogate manner, meaning MHs are not aware of the multicast routing. The BS need to decapsulate downlink packets to MHs, since the traffic is arriving via the multicast tree.

The initial version of MMPv6 was mostly concerned with mimicking the protocol's performance to the IPv4 version of the protocol. Currently, work is under way to enhance the protocol with some IPv6 possibilities: Hop-by-hop headers for combining control messages (CBT and Mobile IP signalling), dynamic/stateless creation of multicast care-of-address, adaptation of some control features already proposed in IPv6 (such as the Registration Request messages used in Hierarchical Mobile IP which contain addresses of intermediate Mobility Agents: in MMP's case this can be utilised to contain the address of the Gateway).

As already indicated, the next Chapter gives results of the original testing of MMP for IPv4 (although in IPv6 the protocol indicates a similar performance [7]). However, the remaining Chapters 4,5 and 7 use a collective approach in analysing IP mobility problems. This means that, unless otherwise stated, the analysis is conducted generically for all issues concerned with both IPv4 and IPv6.

CHAPTER FOUR

Simulation of Multicast for Mobility Protocol

Chapter Overview

This chapter focuses on performance analysis of MMP. The chapter commences by giving an insight into strategies for testing performances of IP mobility protocols and continues with a description of OPNET Modeller, the software tool used for conducting the simulations. Some attention is initially given to explaining the particular simulation setup and selections of parameters used. The simulations and associated results are divided into two categories: handover performance and protocol overhead, intended to extract the two main performance properties. Two more protocols are also tested alongside MMP mainly for comparison purposes: Mobile IP and Hierarchical Mobile IP. The chapter ends with a validation model for proving the correctness of simulations and includes further mathematical analysis of the performances of simulated protocols and the consequent conclusions on the pros and cons of MMP based on the simulated scenarios and other relevant issues.

4.1 Strategies for validating performances of IP mobility protocols

As indicated in the outline of design principles of IP mobility protocols (see section 2.3.2) there are four essential design goals: minimising *handover latencies*, sustaining reasonable protocol overhead, maintaining desirable characteristics of the protocol and achieving compatibility with other IP protocols. Apart from the descriptive analysis there are additional ways of assessing performances of IP mobility protocols:

• Validating performance of a protocol through mathematical modelling: Some operations of a mobility protocol can be represented in mathematical terms. The *handover latency*, for example, can be expressed in terms of the trip times of the update messages considering the layout of the particular network and its characteristics (see section 4.4 and section 4.5). The time needed for an update message to reach the "cross-over" router during a handover can be calculated including all relevant delays (i.e. transmission, processing, propagation…) and performing enough iteration to represent a particular *handover distance*. Protocol overhead can also be represented in mathematical terms, by calculating the cost of a protocol by including all processes involved in the protocol steps and comparing them with similar processes of comparable or rival solutions. However, the characteristics of operations of IP mobility protocols are generally conceivable, i.e. much of the eventual results of an analysis of protocol procedures are evident from the mere explanation of a protocol's mechanisms. Thus, the mathematical modelling would only appear as the abstraction of the descriptive analysis of a certain protocol. This is one of the primary reasons why mathematical modelling has not been applied in the IP research community as the dominant tool for validating performances of mobility protocols.

• Simulation of protocol mechanisms: The whole or parts of an IP mobility protocol can be simulated to observe the performance of the protocol under various conditions of networks, users, traffic… Simulation tools allow great flexibility during the analysis, but, at the same time, require realistic modelling platforms to inject validity into the results obtained due to the dangers of over-assumption, i.e. over simplifying the operation of the tested protocol in order to simulate and hence obtain a particular subset of results.

• Test-bed implementation: Implementing a protocol in a real test-bed with proper IP network components and their elements, such as routing kernels, protocol stacks and physical interfaces, gives more "weight" to the results obtained. The main drawback of test-bed implementation is limited flexibility and scale of the test-beds. Test-beds can rarely include a large number of network elements (routers, MHs, BSs…) due to physical and other limitations. While some IP processes can easily be validated on platforms offered by standard test-beds, the flexibility offered by simulations is rarely matched.

For assessing MMP's performance two methods have been applied and presented in this chapter: simulations and some additional mathematical modelling for validation of simulations and further analysis of the simulation results. Two more protocols are considered: Hierarchical Mobile IP and Mobile IP. The simulation results and scenarios are further analysed by extending the models used for validation of the simulation results. It is considered that this combination can offer a useful insight into the performance-related properties of the three mobility protocols for the considered testing scenarios. This chapter is author's own contribution and the simulation tool used was OPNET Modeller, a dedicated software package allowing realistic modelling of network topologies and the inclusion of custom or user defined internal layered protocol architectures along with facilities for generating and routing different packet flows. The OPNET Modeller presents an excellent compromise between

simulations and test-bed implementations as the tools for evaluating IP mobility protocols. Simulation is concerned with extracting two main performance indicators: handover performance and protocol overhead. Analysis of those two sets of results can further assist in validating MMP's performance relative to the design goals. The chapter ends with an overall conclusion on MMP and other simulated mobility protocols and provides a basis for the direction of further research presented in the next chapters.

4.1.1 Description of OPNET Modeller

OPNET (Optimum Network Performance) is a "commercial software package" produced by MIL3 (Third Millennium Technologies) for modelling and simulating communications networks, protocols and distributed systems. The software package contains extensive libraries of existing protocols (e.g. IP, ATM, TCP,…) which can be adapted to any network setup used in simulations of various systems from Local or Wide Area Networks to mobile radio networks and satellite networks. OPNET allows for flexible specifying of source traffic either by providing external and user built traffic models or by utilising available OPNET traffic models. Behaviours and performances of modelled systems can all be analysed via discrete event simulations. The package incorporates tools for all phases of simulation study including models design, simulations, data collections and data analysis. Developing systems for simulations in OPNET is performed using four *editors* that capture the characteristics of a modelled system's behaviour. The four *editors* and their main uses are (see Figure 4.1):

- Network Editor (see Figure 4.2): Used for developing network models, which are made up of subnets and node models (e.g. routers, base stations, access points,…) and connecting links.

- Node Editor (see Figure 4.3): Used for developing node models, which are objects in the network editor models. Node models are made up of modules with process models (additionally transmitter/receiver elements, queues structures…), and may also include parameter models.
- Process Editors: Used for developing process models, which control module behaviour and may include parameter models (typically made in C/C++ codes defining the key aspects of node functionality).
- Parameter Editor: Used for developing parameter models, which are profiles that characterise key aspects of the system (this may include link properties, packet structures,…).

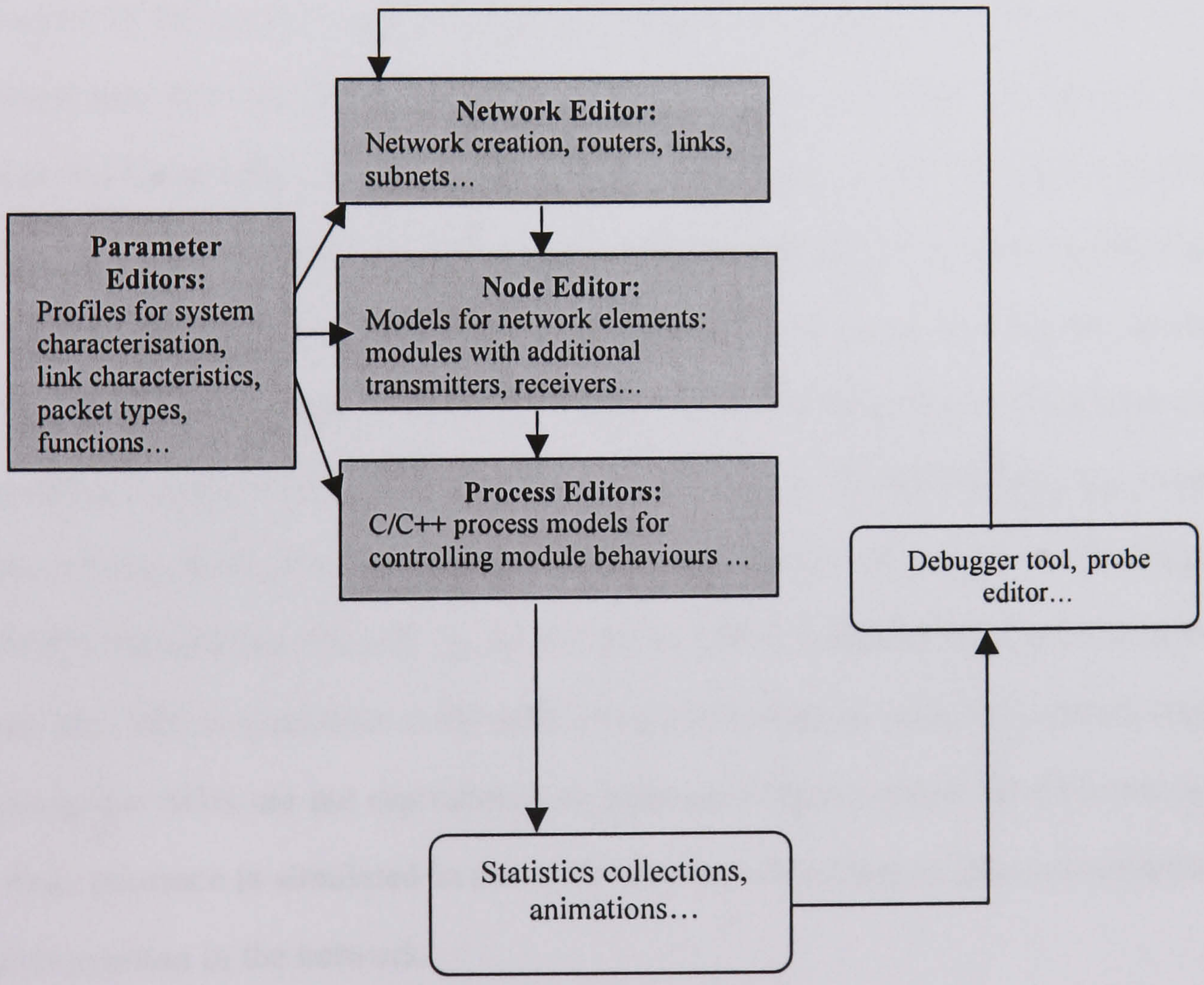


Figure 4.1. Key OPNET functional blocks and their interactions

Finally, systems simulated using the four editors are examined and utilised via statistics collections for post simulation analysis, interactive control of simulation processes (debugger) and animations (for generating default or custom-defined animations of modelled systems).

4.1.2 Description of the Simulation Setup

Topology of the network applied in the conducted simulations is identical to the one shown in Figure 3.5 in Chapter 3. OPNET representation of this network is presented in Figure 4.2 taken from the Network Editor showing all routers, BSs and their interconnections. The foreign network domain (micro mobility domain) contains a collection of full duplex point-to-point links for interconnection of IP routers and four Ethernet-type bus links for connecting "IP-capable" BSs to local LAN routers (Local routers in Figure 3.5). The global Internet environment consists of a HA, a CH and a foreign network Gateway, all mutually connected through full duplex point-to-point links. There are 12 wireless cells in the network each served by a single BS. In Figure 3.5 cell 1 is served by BS 1 or BS 8 in Figure 4.2. In all simulations, cells form a one-dimensional structure where the individual cell width is 30 meters. MHs are assumed to move in a one-dimensional line where the cell width corresponds to the length of the MH's trajectory in the cell. As shown in the OPNET Node Editor level diagram in Figure 4.3, MH is simulated in the *MHprocessor* module in each BS network element meaning that MHs are not represented as separate components of the Network Editor but their presence is simulated in the *MHprocessor* according to their movements and current position in the network.

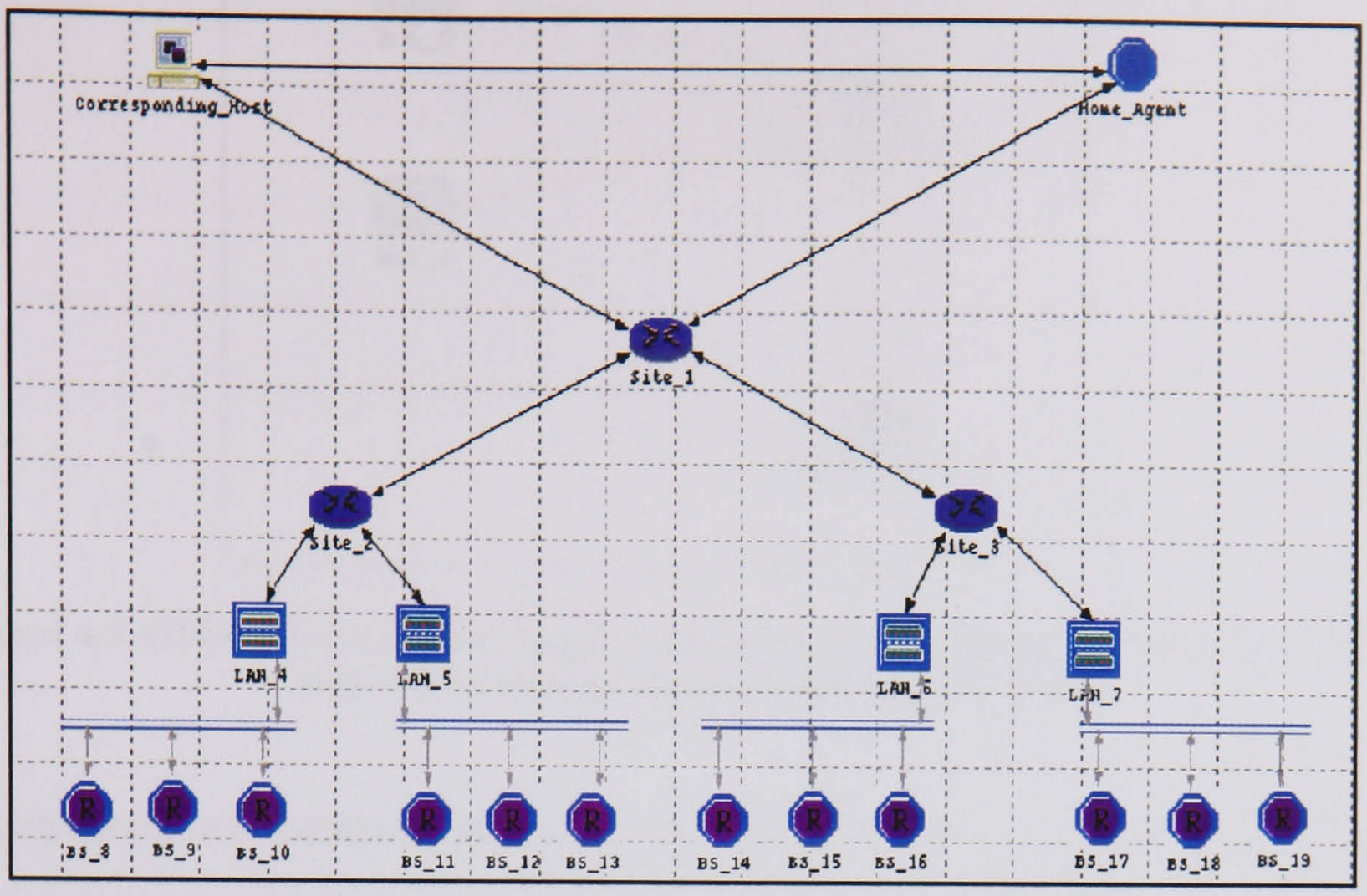


Figure 4.2. OPNET Network Editor image of the network setup used in the simulations

Adjacent cells overlap and a MH is assumed to complete registration with the new BS at the exact time of its disconnection from the old one. In practical terms, this means that as soon as the MH disconnects from the old BS, the new BS relays the relevant control message (for MMP: Registration Requests and CBT Join Requests and for Mobile IP and Hierarchical Mobile IP: Registration Requests) thus intentionally emphasising network layer behaviours. In addition, this strategy is also aiming to eliminate any dependency on specific wireless link technology which have different properties thus transmission rates and delays and are often variable and time-dependent. However, performance of MMP in the *wireless access* section is relative to the investigation since all relevant messages are simulated but is assumed to complete instantaneously. As discussed in the remainder of this section the particular parameters chosen for the simulations of MMP and other tested protocols are a reflection of similar parameters used in the outside work. Some of these external efforts include wireless link properties. This is reflected in the particular values chosen for the network topology and the larger number of routers and links (i.e. hops) in the network compared to other testing platforms.

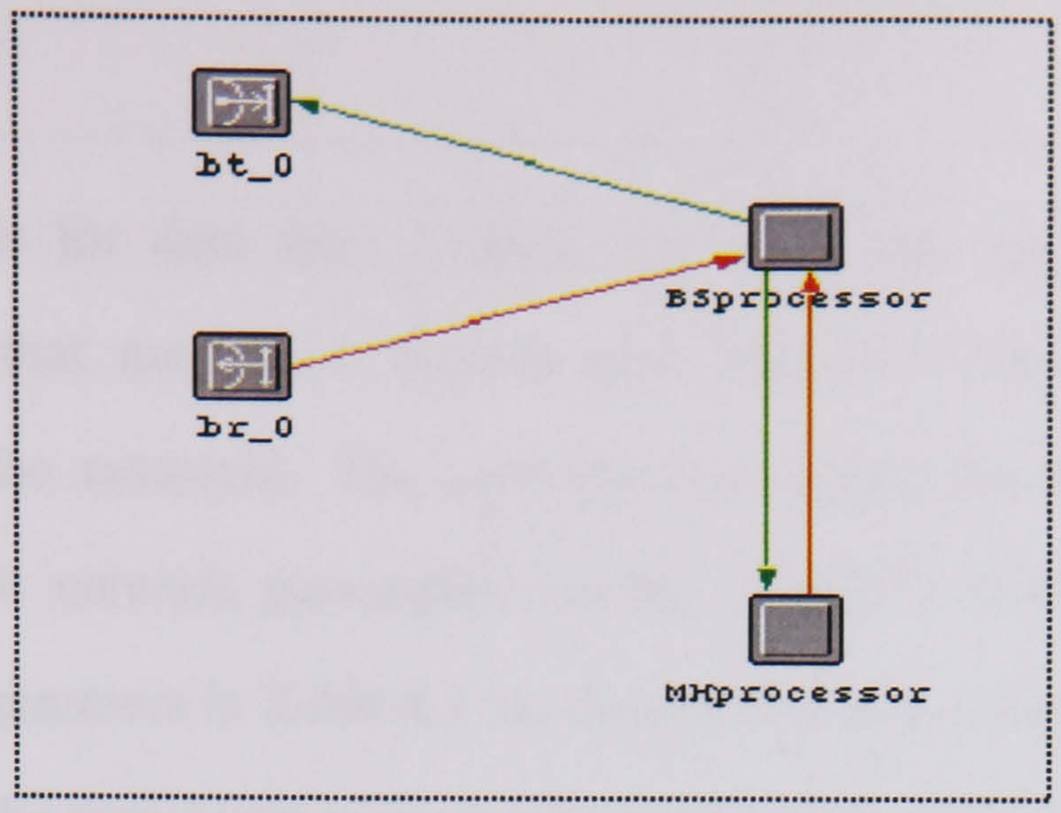


Figure 4.3. OPNET Node Editor image of the modules (processors) and transmitters and receivers in Base Stations of the Network Editor

Two examples of network scenarios are used in the simulations named *high-bandwidth* and *low-bandwidth* networks, both intended to highlight different network scenarios with different links transmission rates and link delays. The values for transmission rates and delays in the Internet links and foreign network links are shown in Table 4.1 for *high-bandwidth* and *low-bandwidth* networks.

Transmission rate delays are applied to all links and routers/BSs. Links are not loaded by any other traffic sources. Transmission rate delays are a standard feature of OPNET network models and are enabled for all simulations conduced. Transmission rate delay is a result of the division of packet's size and the transmission rate of the link. Propagation delays (length of the physical medium over the speed of packets through it) are also included, but are negligible and replaced by the link delays introduced. More on this is given in section 4.4.

	High-Bandwidth Network	***Low-Bandwidth Network***
Internet Links; **transmission rates** (CH, HA, Gateway)	30 Mbits/s	7 Mbits/s
Foreign Network links: **transmission rates** (micro-mobility domain)	10 Mbits/s	2.5 Mbits/s
Internet Links: **delays**	40 ms	40 ms
Foreign Network point-to-point links: **delays**	1.5 ms	3 ms
Bus links **delays**	≥ propagation delay	≥ propagation delays

Table 4.1. Characteristics of the two types of simulated networks: *high-bandwidth* and *low-bandwidth* case

The actual values for data rates in both network cases are an approximation of possible values that may exist in real networks (see chapter 1 for scope and applicability of the research). The intention is to show the impact of the relative differences of the network parameters on the protocol's performance. The values chosen for the parameters in Table 4.1 are also considered to match some real network scenarios where the tested protocols may be deployed as analysed below. Some of the performance metrics such as the concept and implication of *handover distances* in network topologies and hops traversed by protocol messages can be used for deducing protocols' behaviours in networks with different parameters (see section 4.5 and 4.6). An observation and comparison with similar outside efforts can help in understanding the chosen values for the two network cases(*high bandwidth* and *low bandwidth*):

- **Delays**: In simulations of HAWAII [82], a similar test network uses HA, CH and the micro mobility (foreign network) domain. Delays in the Internet links take values of 50 ms and in the micro mobility domain they are set to 5 ms. On the other hand a similar setup was applied for testing Cellular IP and HAWAII in [83] with only 2 ms delays in all links in the network. In the test-bed evaluation of Cellular IP [84] a small size network is used with few routers and BSs but with artificially injected routing delays between MH and CH (and micro mobility domain Gateways), which range from 5 ms to 50 ms. In small scale simulations used for testing IP mobility protocols in [85] the end-to-end delay (CH to MH) was varied from 0 to 100 ms. In simulations of the multicast solution for solving mobility MSM-IP [39] analysed in Chapter 3 a small scale test-bed was used with artificially injected end-to-end (CH to MH) delays of 100 ms. Fast and Scalable Handoffs [34] uses a test-bed with maximum end-to-end delay between end-nodes of 7 ms where each node is connected to BSs connected to a separate Ethernet bus link which are interconnected with another bus link.

- **Transmission Rates**: Fast and Scalable Handoffs [34] uses a test-bed with 10 Mbits/s wired- Ethernet bus links.Fast Handoff Scheme [40] uses simulations with point-to-point links with transmission rate of 10Mbits/s. In the simulation testing of HAWAII [82] transmission rates vary from 10Mbit/s to 155 Mbits/s where the higher transmission rates are applied to Internet links, which are also loaded with background traffic (this is avoided in the testing of MMP thus lower rate are used). In [83] using a similar topology all links have the same transmission rate of 10 Mbits/s. A small-scale test bed is used in [84] with transmission rates of 100Mbits/s and 2 Mbits/s over the wired and wireless links respectively. In a small-scale simulation setup in [85], transmission rates are fixed at 10 Mbits/s. In the testing of multicast for solving mobility MSM-IP [39] analysed in Chapter 3, a test bed was used with transmission rates of 10 Mbits/s for Ethernet bus links. Daedalus [38] uses a test-bed with 2 Mbits/s Ethernet bus links. Note: in cases of high transmission rates used in the simulations the network is typical loaded with background traffic.

Some of the referenced outside efforts, which coincided with development of MMP [1][2] and were relevant at the time of the creation of the simulation platform for MMP are: MSM-IP [39], Daedalus [38] (also discussed in section 3.1.1 as they present attempts to incorporate multicast and mobility), Fast Handoff Scheme [40] and Fast and Scalable Handoffs [34]. The testing environments used in these protocols include bus links (apart from [40] which uses point-to-point links only) and generally have transmission rates of 10 Mbits/s (apart from Daedalus [38] which uses 2 Mbits/s bus links) (the above provides a good match with the simulated topology and in particular the transmission rates of the *high-bandwidth* network).

4.1.2.1 Simulations of Hierarchical Mobile IP[1]

As explained in the second chapter (see section 2.3.3.1), Hierarchical Mobile IP[2] is the basic example of *Proxy-Agent Architecture* and provides a good reference point for comparisons with MMP. The basic steps used in Hierarchical Mobile IP can be extended to represent the general operation of *proxy-agent architectures*. The initial definition of Hierarchical Mobile IP defines a hierarchy of FAs, which are used for localising Registration Requests and hence Replies. The concept of a FA is the default one as specified by Mobile IP and commonly results in a non-optimal shape of the hierarchy of FAs, which are usually the edge nodes in a network. By analogy with the setup used in the simulation of Mobile IP, FAs are BSs, which, if applied to the Hierarchical Mobile IP case, would mean a hierarchy of FAs located at non-optimal locations. This shortcoming has necessitated flexibility in the placement of Mobility Agents (i.e. Proxy Agents), used by some of the more recent *proxy-agent architectures*. The result of the effort is the freedom to place the Proxy Agent at an arbitrary location in the network. The set-up of Hierarchical Mobile IP in this document allows a free placement of FAs.

Figure 4.4 shows the model used in the simulation of Hierarchical Mobile IP with three levels in the hierarchy of Proxy Agents. The top level includes the **top-level FA** being the Gateway for the network, the intermediate level includes **regional FAs** which control bus links and the lowest level includes **local FAs** which are the original Mobile IP FAs having a link layer connection with MHs, or, in this case, BSs with a wireless connectivity with MHs. In contrast to Mobile IP and MMP, which use the simplest form of Agent Advertisement messages, embedded in the wireless link beacons, Hierarchical Mobile IP uses an extended version of the message since it

---

[1] Note: This section presents a generalisation of the mechanisms of Hierarchical Mobile IP intended to extract the main performance properties of the protocol and provide relevant comparison platform for the simulations. The actual specification of the protocol may vary over time but it is assumed that the simulated properties will remain the same.

[2] Actual the Hierarchical Mobile IP examined in this section is similar to Regional Registration [30] but the term Hierarchical was used because it is self-explanatory.

always includes the hierarchical "tree" of FAs in the message. This increases the overhead in the wireless part of the network. However, since all simulations are focused on comparing network layer characteristics of the protocols, the increased overhead of the larger Agent Advertisement messages does not influence the results and comparisons.

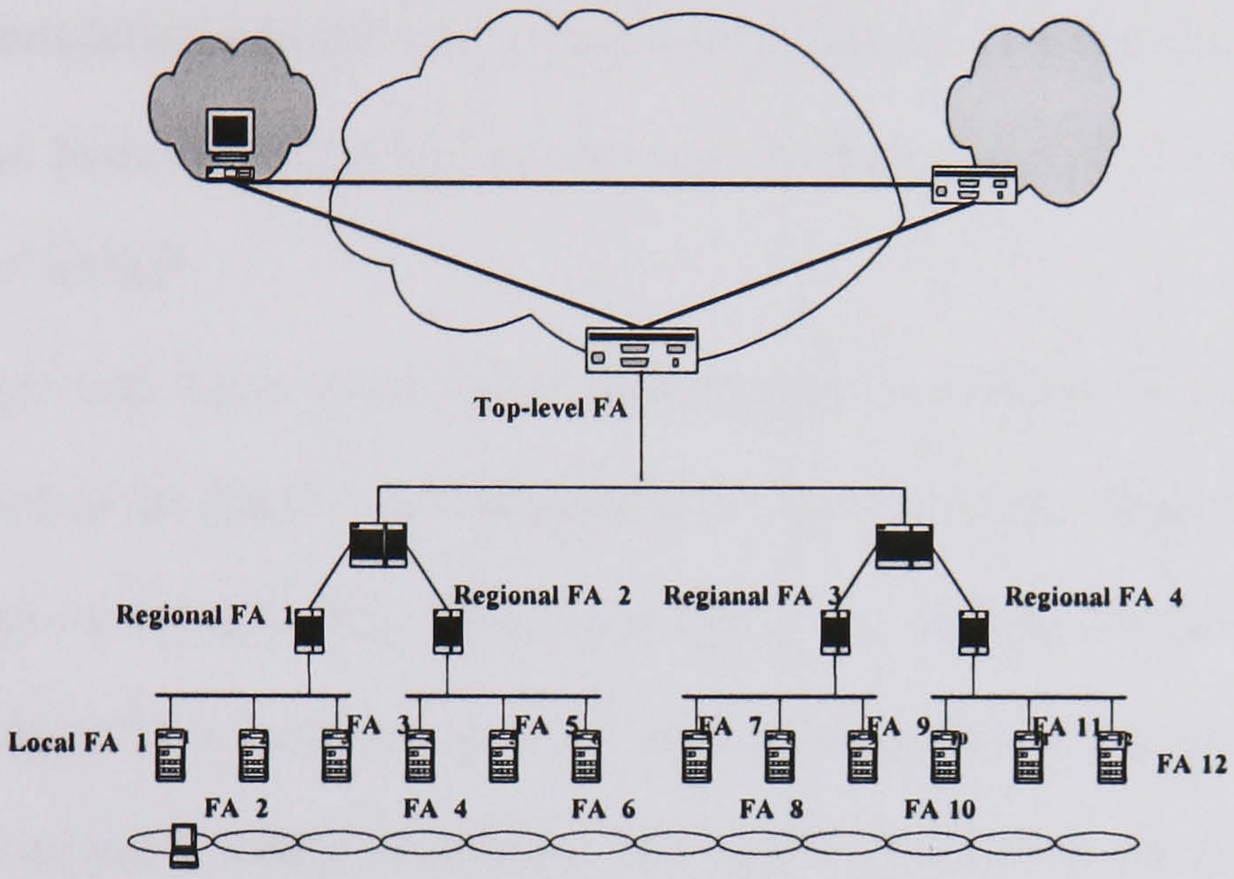


Figure 4.4. Hierarchical Mobile IP setup

Handover procedure in Hierarchical Mobile IP includes decision making by MHs regarding where to send the Registration Requests once the MH has entered the new cell. The essence is in determining the anchor FA, that is, the "cross-over" FA of the old and the new "tree" of FAs. For example a MH handing over from FA 3 to FA 4 requires it to send the new Registration Request to the top-level FA and not to HA since the top-level FA was on the previous tree including HA, top-level FA, regional FA 1 and local FA 3 whereas the new tree includes HA, top-level FA, regional FA 2 and local FA 4. Another important property is that FAs are not aware of what is happening beyond the first FA down the hierarchy because during the transfers of Registration Requests, the source address is replaced with the address of the current FA.

4.2 Handover Performance Simulations

Handover simulations are intended to reveal transfers of control messages during the relevant stages of handover execution and to observe the impact of those protocol procedures on the overall service to MHs, in particular delivery of packets to MHs. Additionally, simulation results can be used to justify the overall protocol operation or, at least some features of it, such as the use of the new *MMP Instruct* message in the simulation of MMP.

The OPNET network setup used in all simulations is shown in Figure 4.2, and is identical to the one in Figure 3.5 discussed in the previous chapter describing the protocol features of MMP. The simulation setup for testing handover performance consists of a single MH, initially in cell 1 and handing over to adjacent cells until it reaches cell 12 where it stops reception. The speed of MHs is set to 2 *m/s*. The test network is accordingly modified to emphasise the network layer performance of the tested protocols, hence other parts of the system, such as packet flows from MH to CH (i.e. the uplink traffic) and wireless link effects, are deliberately omitted. Similar strategy, regarding the uplink traffic, is also applied in other simulations and test-bed validations of IP protocols [82][83][84][85][39] where the particular simulations of HAWAII in [82] claim that the uplink traffic achieves similar performance. This approach is consistent with the analysis of mobility protocols presented in section 2.3.2 where the Handover Execution Delay is assumed to be an independent characteristic of all mobility protocols and its investigation needs to be performed with an awareness of the operation of lower layers. More analysis of the movement detection procedures and handovers is presented in Chapter 6. Additionally, MMP does not include any novelty in the design of protocol procedures in the *wireless access* section, hence does not present the main target in the simulation process.

The *advance registration* protocol mechanism explained in the previous chapter is intentionally avoided since it would produce additional performance enhancement during handovers and the network layer performance of MMP would be less apparent. Additionally, in order to execute the *advance registration* feature, there needs to be a supporting model in which a MH uses the wireless link. This would require the simulation of a model for achieving simultaneous connections to both the old and the new BS during handovers in order to form the routing tree prior to the final change of the serving BS. Since, as already mentioned, obtaining network layer performance is the priority of MMP simulations, any extra complexity in the *wireless access* section is avoided and it specifically discussed in Chapter 6.

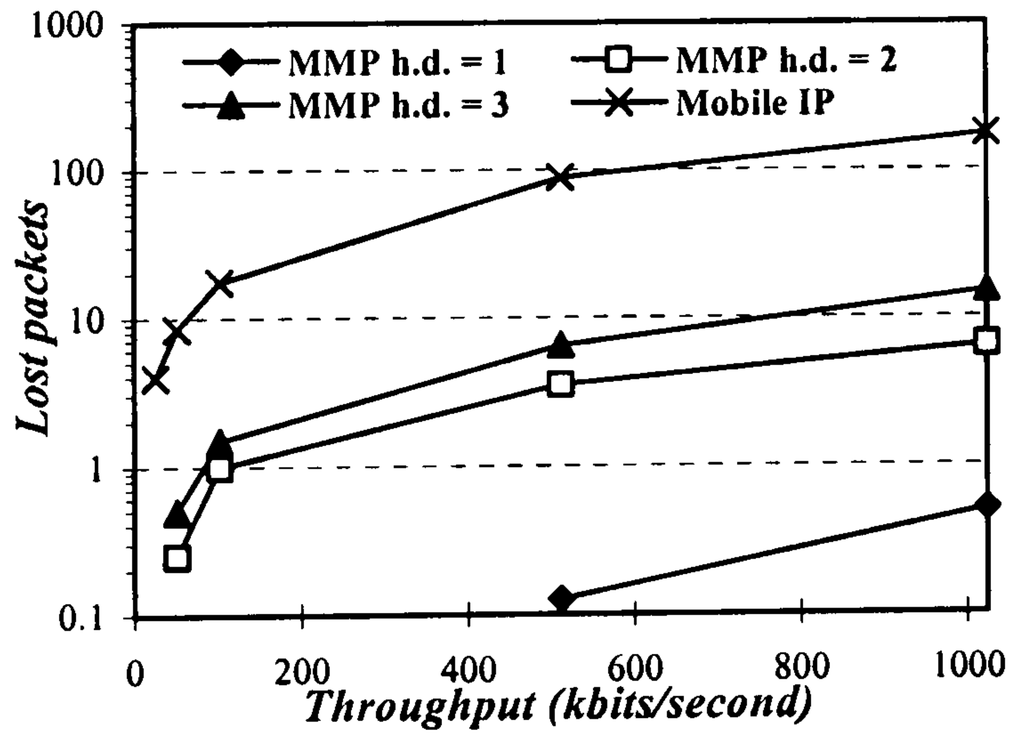
Mobile IP is simulated in accordance with the IETF design, explained in chapter 2, where all BSs are configured to act as FAs.

UDP traffic was used as the traffic load generated by the CH with offered throughputs ranging from 25.6 kbits/s to 1.024 Mbits/s[3] for the fixed packet size of 64 bytes. Two traffic models are used: **constant** and **exponential**. Constant and exponential traffic models signify two types of generated packet streams where the terms constant and exponential refer to the probability distribution of the interarrival times of the packets generated. Sizes of control packets are set according to MMP/CBT and Mobile IP specifications.

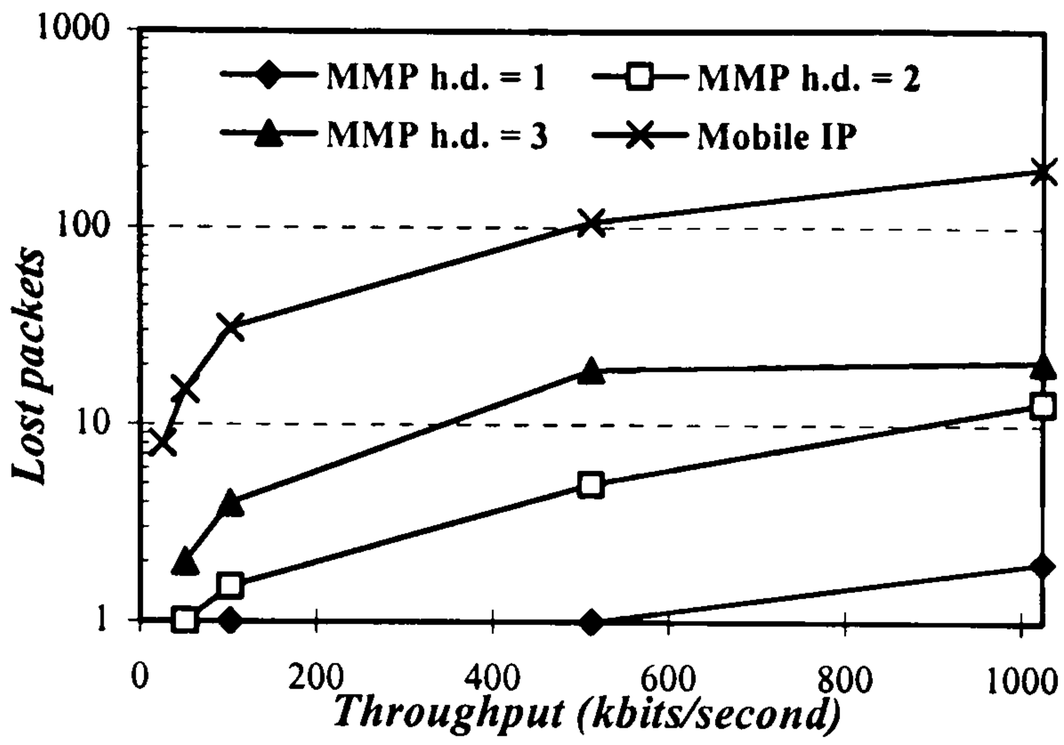
All the results are obtained after extensive simulations and for each set of results twenty simulation runs were performed with different random sequences to level out possible deviations. Results for both simulated networks for both MMP and Mobile IP are shown in Graph 4.1, Graph 4.2, Graph 4.3 and Graph 4.4. MMP's results are shown separately for all three cases of *handover distances*. Results verify that the *handover* latency, and thus the packet loss, is mainly influenced by the path taken by the update messages (Join Request in MMP, Registration Request in Mobile IP)

[3] HAWAII [82] testing uses UDP traffic of 65 kbit/s, in [83] UPD traffic rate is set to around 16 kbits/s, in [84] 40 kbits/s.

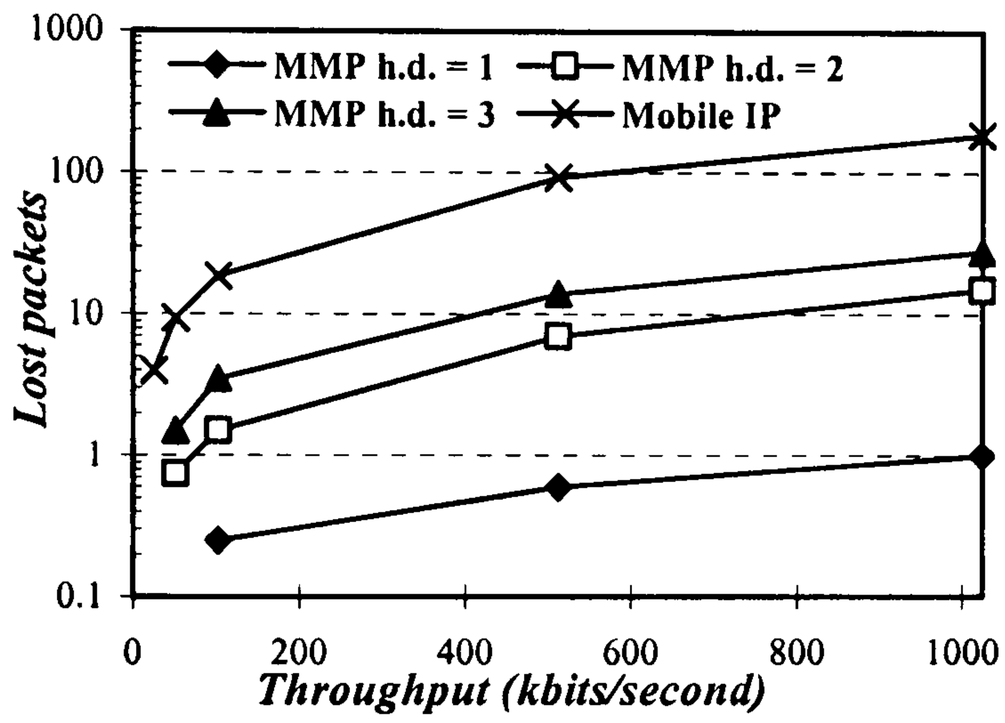
before the packet flow is diverted to the new BS. As the update messages traverse relevant parts of the network they "absorb" the link delays introduced and additionally, all inbuilt OPNET delays. This significantly influences Mobile IP handovers due to the large delay in the link between the Gateway and the HA, while in MMP, path updates are more localised, especially for small *handover distances*. Packet losses are significant during handovers in the case of Mobile IP while MMP behaves efficiently even in the worst scenarios when the *handover distance* is three. From the graphs: in *high-bandwidth* networks when traffic load is 1020 kbits/s and *handover distance* (h.d.) = 3, 2 and 1, the difference between the performances of Mobile IP and MMP is 158.5, 167, and 173 lost packets respectively for the constant traffic models and 180, 188 and 199 lost packets for the exponential traffic model, all in favour of MMP (this also indicates the differences between different *handover distances* in MMP). Section 4.4 includes method for validations of simulation results, which are expanded and shown in section 4.5 for some further performance analysis based on the simulation scenarios.



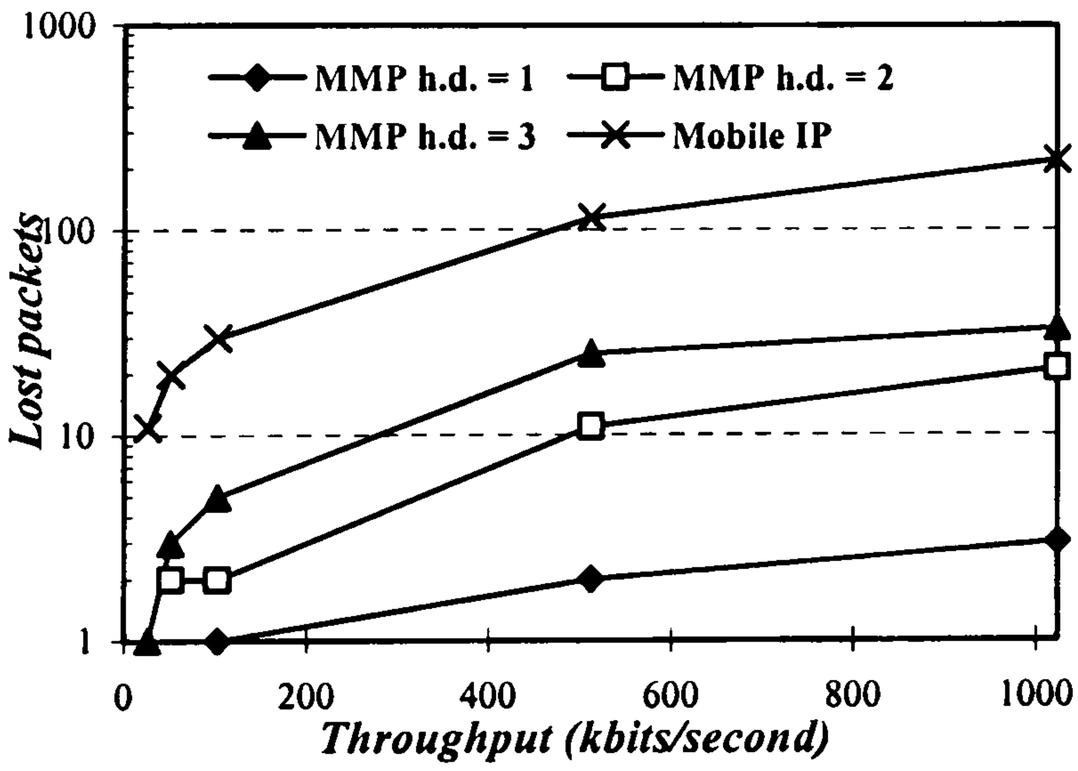
Graph 4.1. Lost packets for the *high-bandwidth* network: constant traffic model, packets size 64 bytes, average of 20 runs.



Graph 4.2. Lost packets for the *high-bandwidth* network: exponential traffic model, packets size 64 bytes, maximum values.



Graph 4.3. Lost packets for the *low-bandwidth* network: constant traffic model, packets size 64 bytes, average of 20 runs.



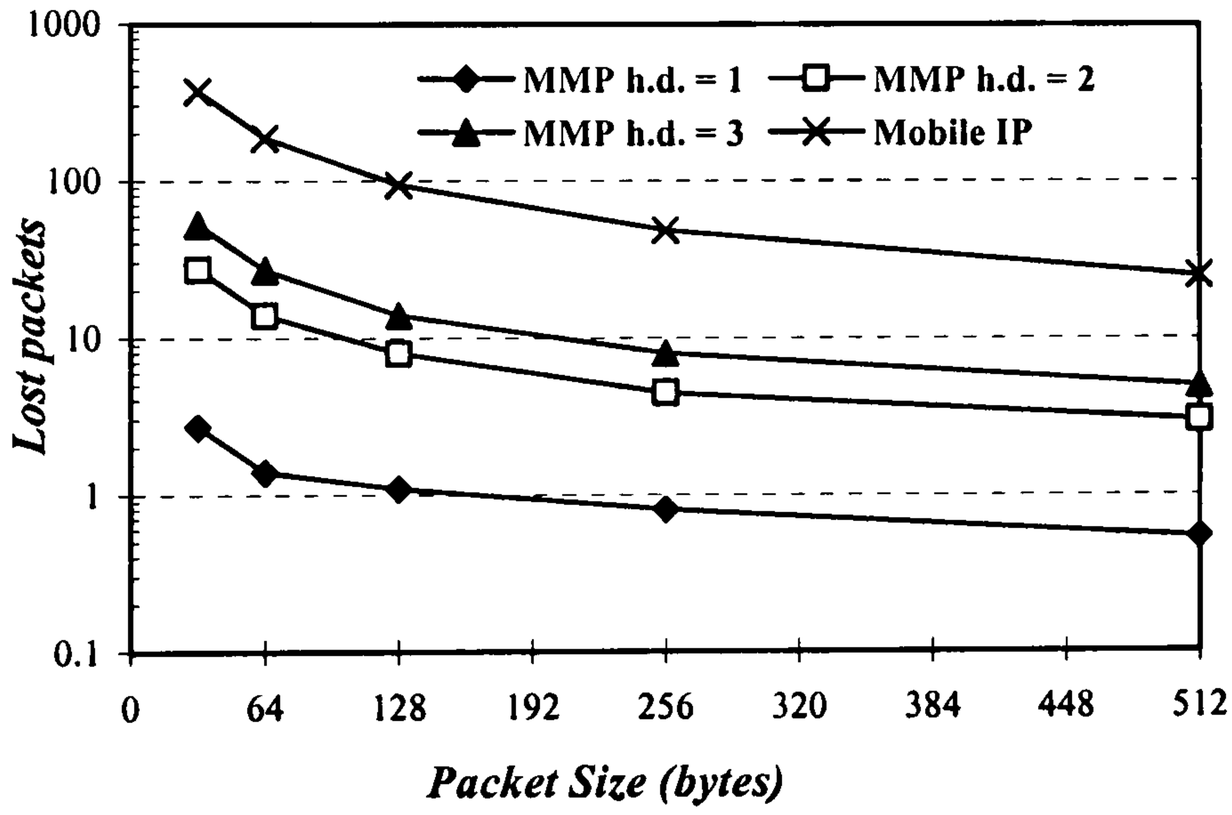
Graph 4.4. Lost packets for the *low-bandwidth* network: exponential traffic model, packets size 64 bytes, maximum values.

From the graphs: in *low-bandwidth* networks when traffic load is 1.020 Mbits/s and h.d.=3, h.d.=2 and h.d.=1, the difference between Mobile IP and MMP is 158.6, 171.6 and 185.6 lost packets respectively for the constant traffic models and 186, 198 and 216 lost packets for the exponential traffic model, all in favour of MMP. The slight increase in the number of lost packets (for the same traffic pattern) in the case of the *low-bandwidth* network occurs due to the "slower" updating process, that is, the update messages (Join Requests and Registration Requests) requires more time to reach the relevant "anchor" nodes when the transmission speeds are lower and link delays are increased (the "anchor" for Mobile IP is the HA, for MMP in Figure 3.5 for h.d.=1 Local routers, for h.d.=2 Site routers and for h.d.=3 the Border Router, i.e. the "cross over" routers). As far as the packet flow is concerned, effects of different transmission speeds and link delays are not the main reason for the increase in *handover latencies*. Actually, a packet stream only appears "shifted" by the *low-bandwidth* transmission offset (summation of all link and transmission delays relative to the *high-bandwidth* case) but not dispersed. This is especially true because all transmission rates are higher that the transmission rate of the traffic(offered throughput), thus no congestion occurs (as there is no other traffic in the system).

The average packet losses in the case of the exponential traffic model are the same as in the constant traffic case. Hence, for the exponential traffic model, the extreme cases are shown when the observed packet loss was at the peak (maximum) value. Additionally, comparing the extreme cases for the exponential traffic model under MMP and the average packet loss under Mobile IP for the constant traffic model, MMP still offers a significantly faster route updating (eg. for h.d.=3 , h.d.=2 and h.d.=1 the difference is 152.5, 160.5 and 171.5 lost packets for the *high-bandwidth* network and 153.6, 165.6 and 183.6 for the *low-bandwidth* network for a traffic load of 1024 kbits/s).

It was observed that some of the results for the constant traffic model are not always the same for every simulation run with the same input parameters (hence the non-

integer values for some packet losses). These variations are not caused by possible deviations due to the random nature of the simulations. In fact, they happen because of the occasional occurrence of a "trapped" packet between the handover "cross over" router and the old BS during the handover. In other terms, a packet can flow through the old tree branch (from "cross over" router to old BS) while the updating process from the new BS is ongoing. In the case of the most successful handover, when *handover distance* is one, packets are also occasionally "trapped" resulting in a lost packet where, in theory, there should be no packet losses since the effective *handover latency* is smaller than the interarrival period of packets. This situation is most obvious in the MMP results for small *handover distances* where the final value displayed in the graphs is sometimes less than one, due to the occurrence of an occasional "trapped" packet (see Graph 4.5).

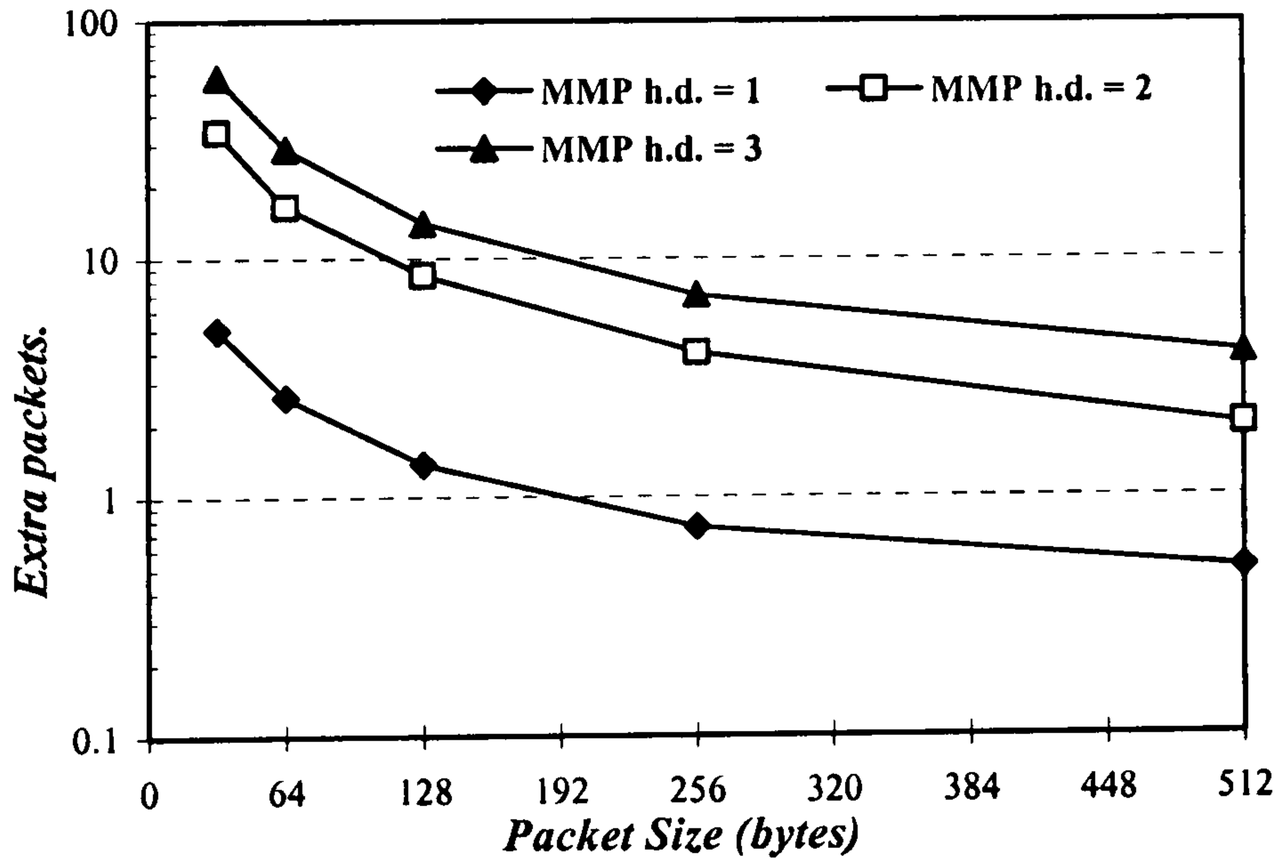


Graph 4.5. Lost packets for low-bandwidth network: constant traffic model, offered throughput 1.024Mbits/s, average of 20 runs.

In Graph 4.5, the effects of varying packet sizes for the *low-bandwidth* network are shown where the offered throughput was fixed at 1024 kbits/s for the constant traffic generation model. Due to the smaller interarrival times of smaller packets, the packet stream appears "denser" and hence the losses during route updates are greater, i.e. the period of *handover latency* "consumes" more packets when there are more of them

(because the byte throughput is fixed, when packets are smaller, more packets are actually sent over an arbitrary interval of time).

The proposed MMP-specific *MMP Instruct* message is also simulated and results are shown in Graph 4.6 where extra packets indicate "wasted" packets on the old routing tree branch before the *MMP Instruct* message reaches the old BS. The number of extra packets can also be interpreted as a time delay but since this is also dependent on the architecture and characteristics of network links, graphical representation seems more appropriate. The largest number of wasted packets occurs when *handover distance* is large because the *MMP Instruct* message needs to traverse six hops before it reaches the old BS (see Figure 3.5). The results range from a case when only one packet gets wasted occasionally for a *handover distance* of one and packet size of 512 bytes, to the extreme case where around 60 packets are wasted for the *handover distance* of three and packet size of 32 bytes. In all cases, the introduction of the new MMP control message is a far more efficient solution than the standard method used in default CBT relying on timeouts, which could be in the order of minutes.



Graph 4.6. Extra packets indicate wasted packets until MMP Instruct is received in the *low-bandwidth* case: constant traffic model, throughput 1.024Mbits/s, average of 20 runs.

Hierarchical Mobile IP handover performance simulation results are not shown in the previous graphs since they can be classified into the two categories already shown for the MMP case. Hierarchical Mobile IP in the setup of Figure 4.4 produces two sets of results for the handover performance as depicted in Figure 4.5:

a) The first one is identical to the results for MMP when *handover distance* is one. This includes handovers between local FA 1 and local FA 2, local FA 2 and local FA 3, local FA 4 and local FA 5, local FA 5 and local FA 6, local FA 7 and local FA 8, local FA 8 and local FA 9, local FA 10 and local FA 11 and local FA 11 and local FA 12 because of the common regional FA for new and old trees of FAs resulting in a single hop needed for Registration Requests to update the necessary entries.

b) The second one is identical to the results for MMP when *handover distance* is three. This includes handovers between local FA 3 and local FA 4, local FA 6 and local FA 7 and local FA 9 and local FA 10 (for all cases both direction of movements are included) because of the new regional FA for the new and old tree of FAs resulting in three hops for the Registration Request which has to update the "cross-over" router, for this case, the top-level FA. Some examples of packet losses for the *handover distance* of 3 are shown in Graph 4.7 and are identical to the results obtained for MMP (*handover distance* of 1 produced occasional packet losses as described above for MMP due to "trapped" packets and is not shown) (further analysis of Hierarchical Mobile IP is shown in section 4.5).

As in the MMP simulations, processing delays are also neglected in Hierarchical Mobile IP simulations. The handover results for Hierarchical Mobile IP are relative to the placement of FAs in this particular scenario (an additional scenario is considred in section 4.5).

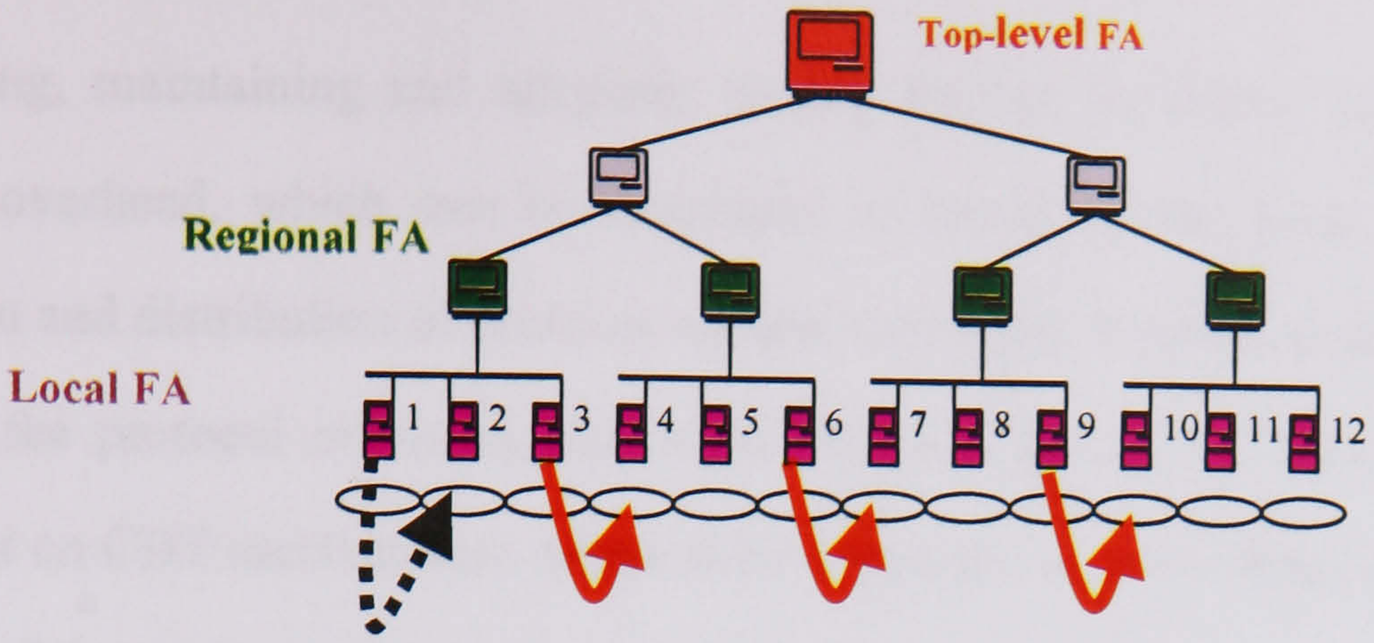
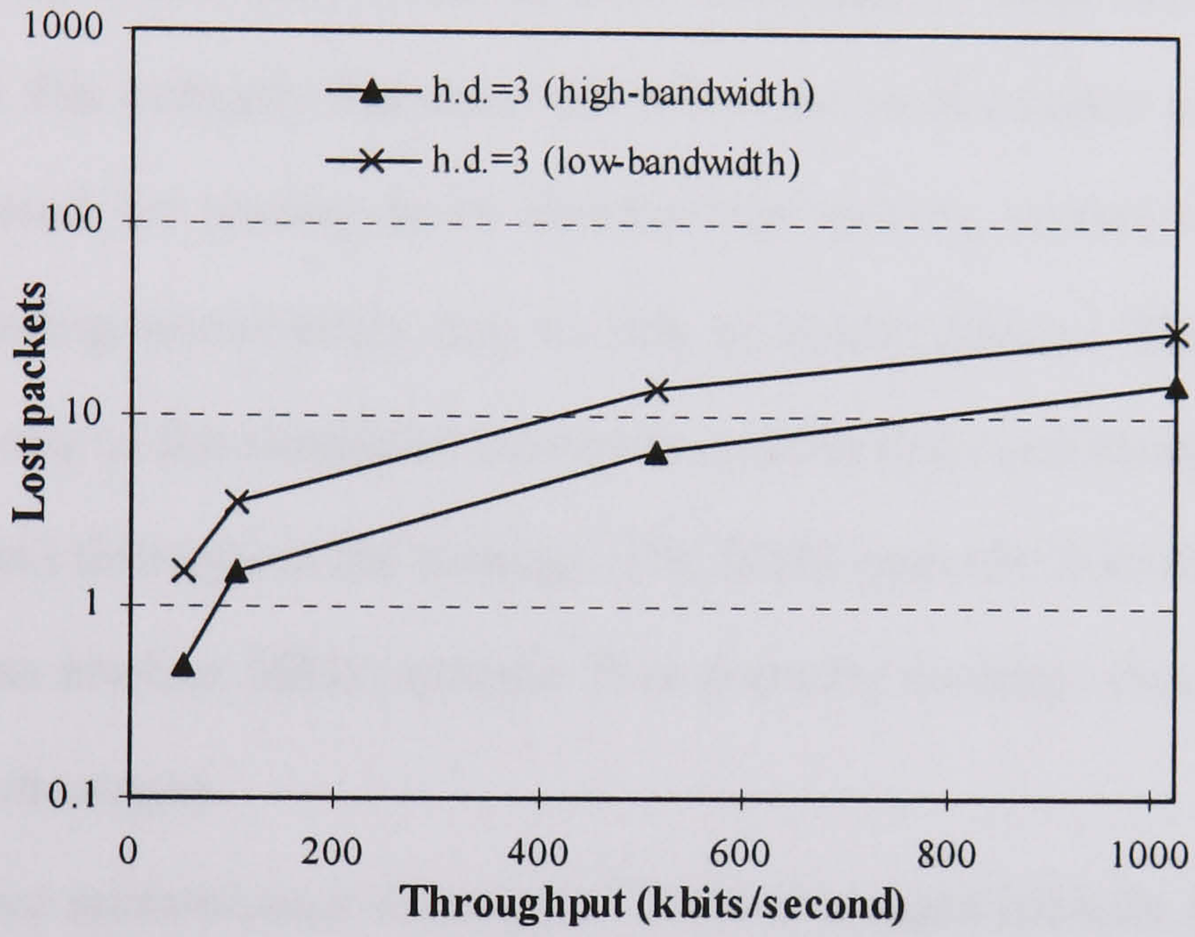


Figure 4.5. Handover cases in the simulated setup of Hierarchical Mobile IP



Graph 4.7. Handover packet losses for *handover distance* of 3 for some cases of H.MIP handovers

4.3 Simulations of Protocol Overhead

Handover latency is one of the key elements for determining the general efficiency of a particular IP mobility protocol. However, satisfactory performance solely considering the *handover latencies* cannot be the only factor for an evaluation of the overall performance of a protocol. As indicated in chapter 2, current research in IP mobility contains various schemes for fast and efficient route updating and location tracking of MHs. Naturally, almost all new schemes, usually starting off as an improvement to Mobile IP, include more complicated routing mechanisms for

establishing, maintaining and adapting routing entries for MHs. This causes more protocol overhead, which can be expressed in terms of the load induced by the generation and distribution of protocol control messages. In order to evaluate MMP in terms of the protocol overhead, the *micro mobility* section of the protocol, mostly dependent on CBT mechanisms, needs to be explained in more detail since it is where the most of the protocol overhead is induced.

There are two types of CBT control messages:

a) *Multicast tree-forming* messages. These messages include *Join Request*, *Join ACK* and *Quit Notification* (additionally, CBT specifies a *Flush Tree* message, which also falls in this category but does not affect the performance of simulated MMP since it is used for tearing down downstream routing entries when an upstream router becoming unreachable due to link or router failure. This situation is not included in any of the simulated scenarios and the links and routers are assumed to be operational throughout the testing). The MMP-specific *Instruct* message can be considered as another MMP-specific *Tree-forming* message since it triggers some of the CBT messages.

b) *Multicast-tree maintenance* messages. These messages include *Echo Request* and *Echo Reply* and are used for maintenance of the "soft state" as explained in section 3.4.3.1. After setting up of the multicast trees in the CBT part of MMP, done by the creation of routing entries, routers are independently sending periodic *Echo Requests* to their next upstream hop towards the Core. Routers, which are receiving *Echo Requests* from their downstream neighbours, create and "send down" *Echo Replies* as acknowledgments. Every router periodically repeats this procedure apart from the Core, which does not have an upstream neighbour and simply replies to received *Echo Requests*.

The exact use of all CBT control messages can be found in the protocol specifications [18]. *Join Request*, *Join Ack* and MMP-specific *Instruct* are explained in the previous chapter. *Quit Notification* is used to start the tearing down of a multicast routing tree

in the upstream direction, that is, towards the Core (Gateway). There are two ways to trigger the transmission of a *Quit Notification* message in MMP:

1. Default mechanisms of CBT: These are timeouts affected by the distribution of *multicast-tree maintenance* messages or general multicast mechanisms. In a native multicast case, BSs would decide on whether to initiate the tree pruning based on the reception of IGMP messages from MHs belonging to multicast groups. In MMP, the analogy of this would be the lack of Registration Requests by MHs. For multicast routers in the network (not local ones like the BS) there is no triggering mechanism for sending Quit Notification since the situation where the child interface becomes null (interfaces to the downstream routers) is not include in simulation. As specified in section 3.4.2, an IGMP *Leave* message (available in the newer versions of IGMP) is not implemented in the *wireless access* section but could provide for a default multicast mechanism to prune the multicast tree from the local multicast routers (BSs in the MMP setup).
2. MMP protocol extensions: This is performed by the explicit pruning of the multicast tree by the end routers. The end routers are BSs and they prune the upstream tree after the reception of an MMP *Instruct* message sent from the new BS during the handovers of MHs.

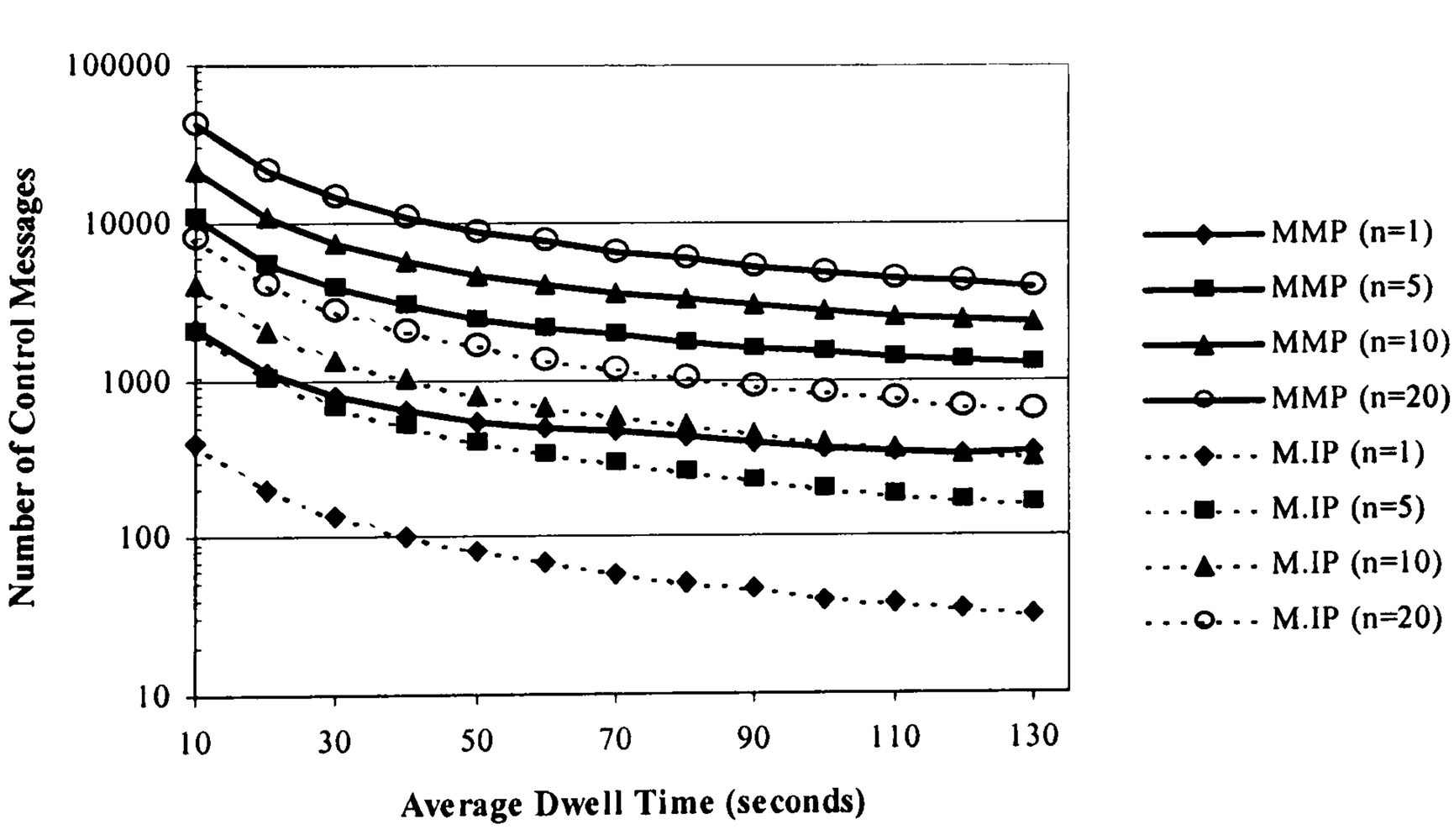
CBT provides some extension for additional protocol procedures when the protocol is implemented over bus links. Standard specifications of protocol mechanisms usually relate to point-to-point connections between routers. Bus link scenarios are extensively included in the MMP simulations as shown in Figure 4.2 and Figure 3.5 where there are four bus links running the Ethernet link-layer access protocol. All four bus links have an identical setup with four multicast routers attached where three of these are the actual local multicast routers for MHs, that is, the BSs. Basically, the need for additional protocol features in broadcast-based mediums such as the bus link comes from the property of such mediums whereby all packets sent on the link are received by link-layer modules of all attached nodes (routers). For unicast

communications in Ethernet driven bus links, packets are usually filtered by the appropriate address resolving mechanisms such as ARP and are hence not forwarded to the network layer. However, the problems in CBT and hence in MMP arise because of the addresses of control packets, which are multicast and sent to an "all-CBT-routers[4]" multicast group address. This means that all routers on the bus link having CBT routing entries, receive all control packets regardless of whether they are intended to only one or a limited subset of routers on the link. In both the native multicast setups and in the particular mobility adaptation of CBT used in MMP, this collective reception of control packets can create problems. Considering a particular MMP scenario:

- A MH is performing a handover to a new BS, where both the old and the new BSs are on the same bus link, for example a handover from BS 1 to BS 2 as shown in Figure 3.5 (of BS 8 to BS 9 in Figure 4.2). In MMP, BS 2 would send an *MMP Instruct* message to BS 1 causing it to send a *Quit Notification* message upstream towards the Core in the desired attempt to "cut off" the old routing branch from the multicast tree of the MHs. The *Quit Notification* is originally intended for an upstream router in the old multicast tree (from the BS 1 to the Core, or to the "cross over" router), which is the Local Router 1 in Figure 3.5 or LAN router 4 in Figure 4.2 sharing the same link with the BS 1. The Local Router does not prune the interface through which the *Quit Notification* arrived from the multicast tree (as would happen in situations with point-to-point links) because BS 2 is still using the same interface. Hence, upon the reception of the *Quit Notification* message (*Quit Notifications* are retransmitted to compensate for potential losses), the Local Router sets up a timer within which a new Join Request should arrive from any possible router still interested in the group membership, that is, the new BS 2. All protocol mechanisms for support of bus links are fully modelled.

[4] For non-multicast supporting links the control messages are unicast to an address determined from the tree-building process. However, for this particular CBT implementation all control messages on the bus

CBT protocol specifications define constants for controlling generations of particular control messages and protocol timeouts. The most important protocol constant relative to the simulation of MMP is MAX_RTX, which indicates the number of retransmissions of some control messages. This particularly applies to the number of *Quit Notifications* transmitted, when a router (i.e. BS) decides to prune itself from the multicast tree. In some scenarios, such as the MMP network setup with a large number of bus links, which cause a complicated pruning procedure (as indicated in the previous paragraph), increasing the number of control messages by increasing the MAX_RTX constant, can lead to a significant increase in the protocol overhead. Another relevant constant is ECHO_INTERVAL, which indicates the period for transmission of *Echo Requests* and logically should be set to the highest possible value to reduce the generation frequency of "keepalive" messages. The default specifications of CBT propose: **MAX_RTX = 3** and **ECHO_INTERVAL = 60.0** seconds.



Graph 4.8. Control messages count: MMP versus Mobile IP (M.IP) (default MMP and 1 Internet Hop), n is the number of MHs.

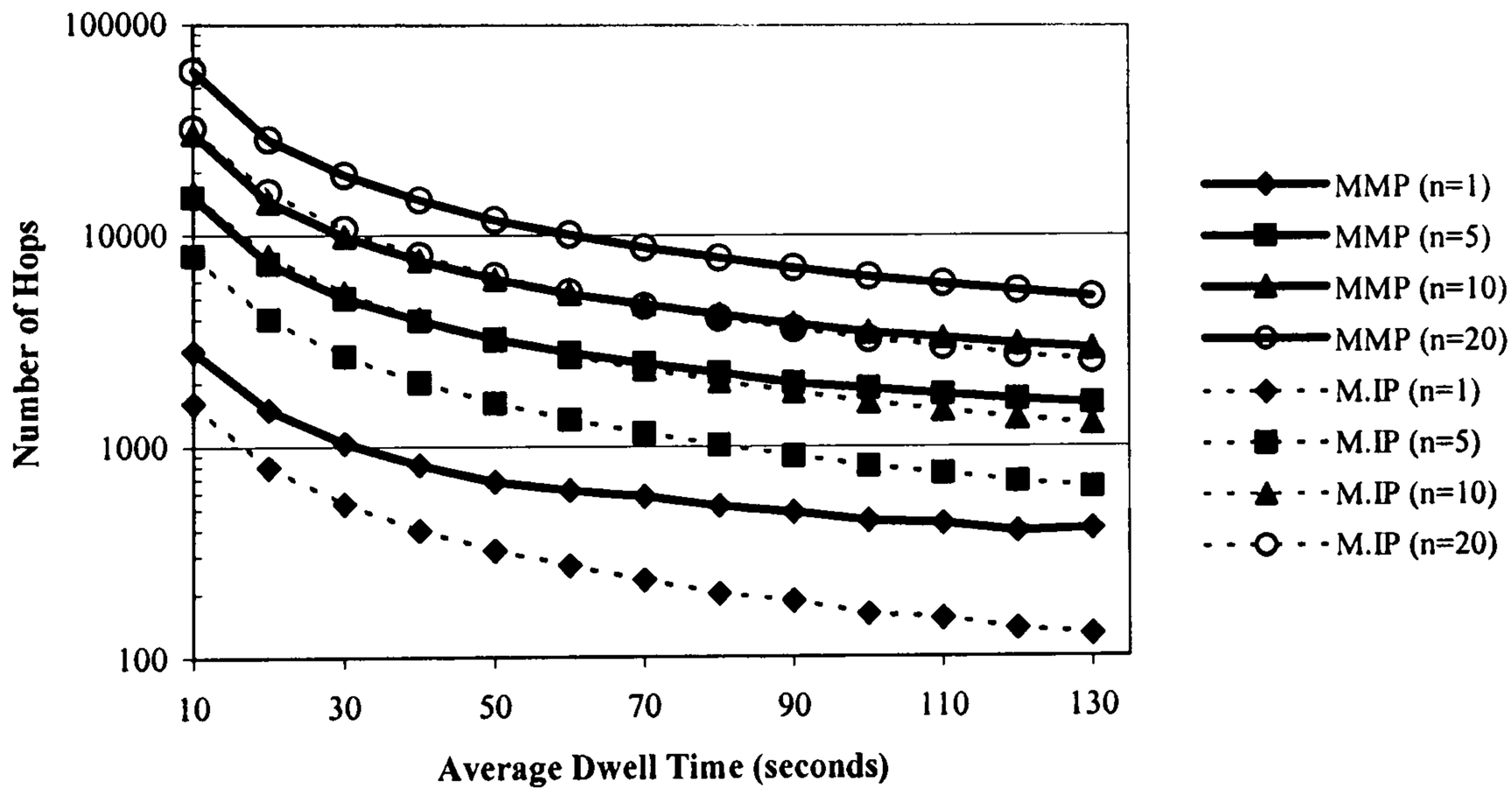
links are multicast according to the CBT specifications.

Simulations are conducted with the aim of obtaining the entire protocol overhead for various combinations of quantity and behaviour of MHs. The network setup for the extraction of the protocol overhead results is the one used in the handover performance testing (see previous section) for the *low-bandwidth* network case. As indicated at the beginning of this chapter and in section 2.3.2, handover performance and protocol overhead have to be weighted again different implementation scenarios to check whether the protocol being tested maintains its initial characteristics in different circumstances. One of those characteristics is scalability. All candidate mobility protocols should scale well in all situations, in particular, when the population of MHs increases in the same Gateway-scoped foreign network domain. For handover performance it can be claimed that the performance would not be greatly affected by an increase in population of MHs, as this does not directly affect the *handover latencies*[5] achieved. On the other hand, for determining and validating protocol overhead, scalability is one of the primary design concerns. Hence, the number of simulated MHs in MMP protocol overhead simulations is varied from the initial case with only one MH in the system to a maximum of twenty MHs. Varying the number of MHs in the system was particularly important in MMP simulations since the increase in the number of control messages generated is not linearly proportional to the number of hosts, particularly because of the MMP/CBT "keepalive" mechanism (explained in the "soft state" section 3.4.3.1) which is independent of the number of multicast groups (i.e. individual MHs) a router is supporting at an instant.

The duration of all simulations is 2000.0 seconds. **Dwell time** is defined as the time a MH spends in a cell and is inversely proportional to the speed of MHs which are moving in a straight line through the one-dimensional cell structure where the length

[5] Although not considered in the simulations, it should be noted that in some situations, extent of the population of MHs could affect the individual performance of handovers. A particular case is when population of MHs is large enough to cause a large traffic flow in the involved network. This situation

of the path though the cells was set to 30.0 *meters*. There a four cases of different populations of MHs: n=1, n=5, n=10, n=20, **n signifying the number of MHs**. For the population case of five MHs (i.e. n=5) the speeds of MHs (hence the dwell times) are distributed in the following way: one MH is moving at the speed corresponding to the **average dwell time** in cells while the remaining four MHs are moving at speeds which are symmetrically distributed: two with higher and two with lower speeds relative to the "average dwell time speed". The overall standard deviation of the higher and the lower speeds is a fraction higher than half the average. The remaining cases of MHs (n=10, n= 20) are set as multiples of the case with five MHs randomly modifying the starting cells so that MHs are evenly distributed across the wireless network at the start of the simulations.



Graph 4.9. Hops count: MMP versus Mobile IP (default MMP and 1 Internet Hop), n is the number of MHs.

A comparison of Mobile IP and MMP is shown in Graph 4.8 considering the number of control messages generated and with MMP simulated according to the default protocol specifications. The logarithmic scale of the results conceals the linear

---

can potentially cause congestion and incur delays in the flow of control messages, in particular, the handover updates, hence causing an increase in the *handover latencies*.

property of Mobile IP results, in other words, the number of control messages generated in Mobile IP is a multiple of the number of MHs in the system. The difference between the two protocols is significant, ranging to around 36000 extra MMP control messages for the fastest moving collection of MHs when n=20 (dwell time =10.0 *s*), to around 300 extra messages for a slowest moving single MH (n=1, dwell time = 130.0 *s*). Due to the limited simulation time of 2000.0 *s*, faster moving hosts (with smaller dwell times) generated more control messages due to executing more handovers. This also indicates that handover procedures are the most important factor influencing the overall protocol overhead. Graph 4.9 shows another important feature for determining protocol overhead: the collective number of hops traversed by the control messages of both MMP and Mobile IP, for each case of population and average dwell time. This includes the hops traversed in the foreign domain as well as the ones "absorbed" in the global Internet. Initially, the Internet environment was set to include only one hop between the Gateway/Core and the HA resulting in an overall single hop for messages flowing between them (Registration Request and Reply for both Mobile IP and MMP). Thus in Graph 4.8 and Graph 4.9 **Internet hops are set to 1** (note: the number of Internet hops does not influence the generation of control messages, only the numbers of hops they traverse).

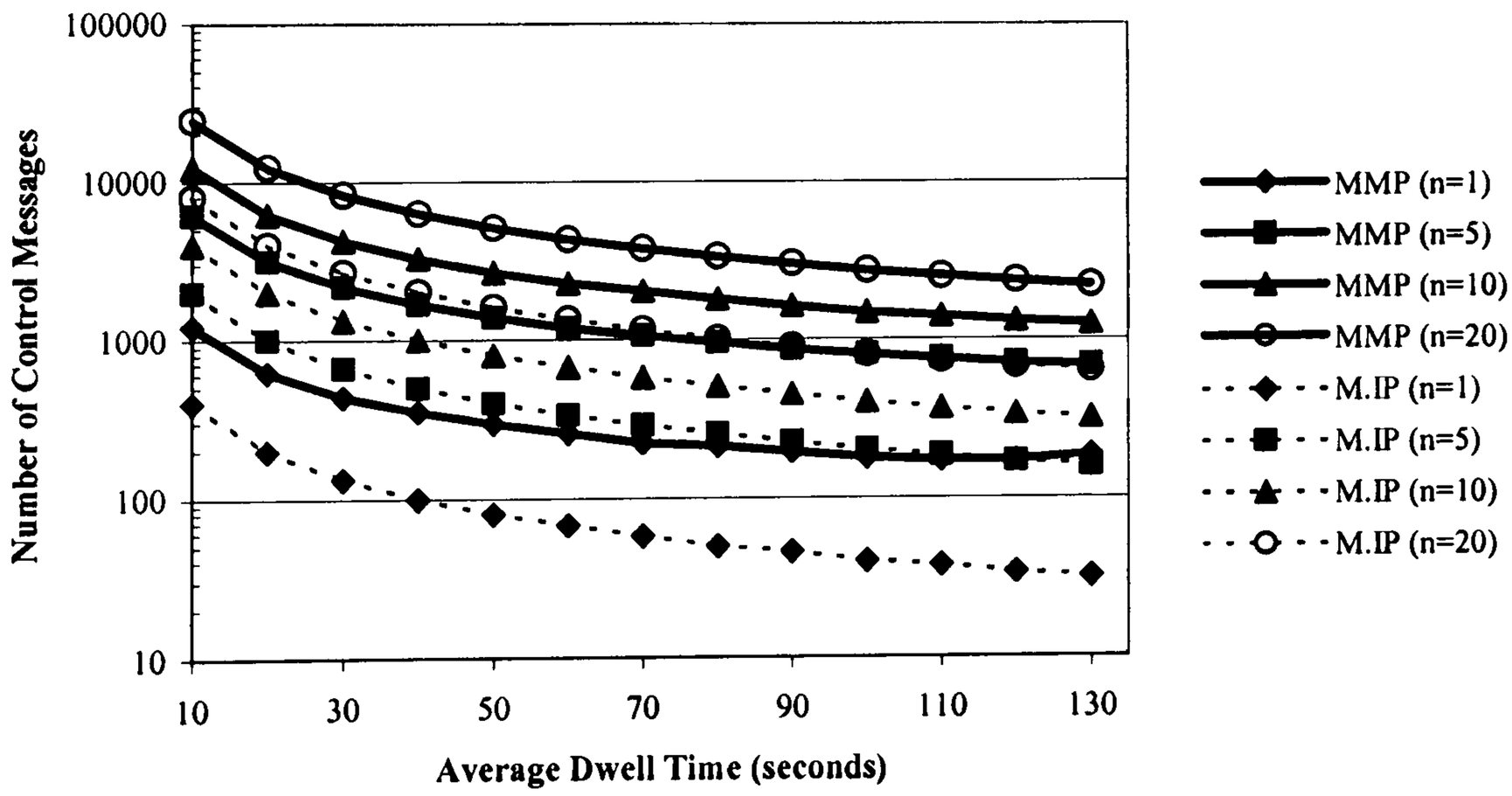
The difference between the two simulated protocols is again significant, ranging from around 28000 extra hops for MMP, for the fastest moving MHs when n=20, to around 270 extra hops for the slowest moving single MH. Graph 4.8 and Graph 4.9 reveal that, unlike Mobile IP, MMP behaviour is not linear especially at instants when the average dwell times are 60 and 120 seconds. This particular behaviour indicates a sudden increase in the number of generated control messages and the hops traversed by them.

The cause of this occurrence is the periodic generation of "keepalive" messages, which are triggered after every 60.0 seconds according to the ECHO_INTERVAL value of the default CBT protocol specifications. Hence, if a MH "dwells" in a cell for

60.0 seconds or longer, this will trigger the execution of the "soft state" mechanisms (Echo "keepalive" mechanisms). This is more relevant in parts of the multicast tree closer to BSs since there is more chance that those sections of the tree contain entries for single MHs, thus directly influencing the execution of the "soft state" since they are the first and only entries of the tree. For tree sections closer to the Gateway/Core it is more probable that they contain entries for more MHs which belong to different "lower branches" (i.e. different BSs) thus having aggregated executions of the "soft state" mechanisms and being less dependent on the movements and dwell times of MHs. The "soft state" triggers for all sections of routing trees are determined at the initial tree creation instant, i.e. when the first entry is created for a MH. Hence, further new entries are neglected, as far as the "keepalive" mechanisms are concerned. Note: it is also likely that "lower sections" of routing trees could contain entries for more than one MH, that is, when there is a multitude of MHs in the same cell. This is more relevant for higher population cases. The relative effects of the increase in the generation of control messages at multiples of ECHO_INTERVALs are more obvious in the case of a single MH because of the sudden generation of "keepalive" messages. For more users, the effects are "smoothed out" because of different speeds and thus of dwell times, which cause the "keepalive" messages to be transmitted at different times. Additionally, relative effects of an increase on the y-axis (number of control messages) are more obvious for lower values of the generated control messages, which occur when there are less MNs.

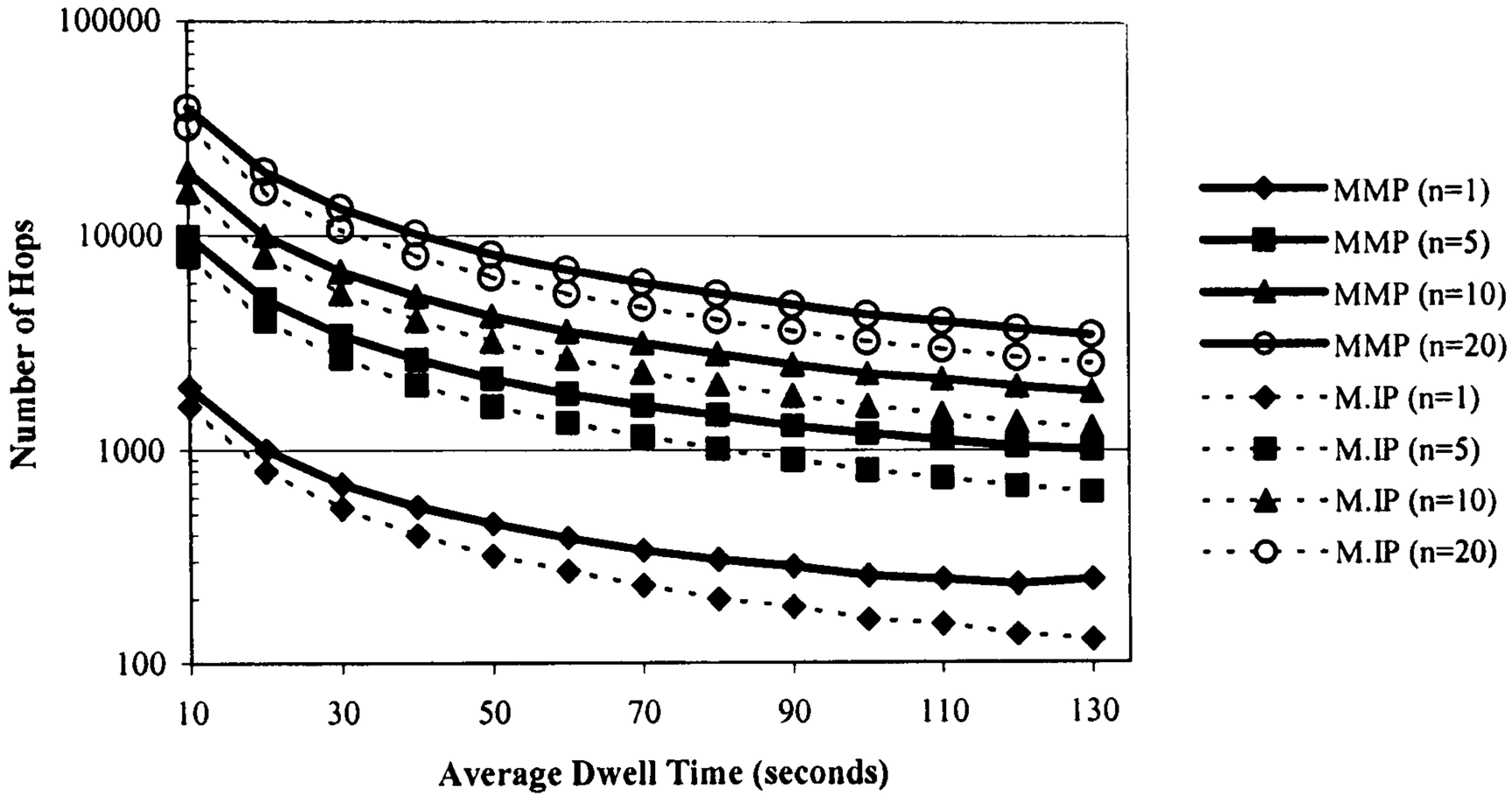
In order to show the effects of varying the values of protocol constants, MMP was modified by setting new values for CBT protocol constants such that MAX_RTX = 1 and ECHO_INTERVAL = 120.0 seconds. This was expected to produce reductions in the protocol overhead for MMP due to the smaller number of messages generated. Graph 4.10 and Graph 4.11 show two comparisons regarding the number of control messages generated and the hops that they traverse respectively (modified MMP means the new values for protocol constants are included). Comparing Graph 4.10 to

Graph 4.8 and Graph 4.11 to Graph 4.9, since Mobile IP results are obviously unchanged, MMP results seem to be "shifted down" because of a reduction in the protocol overhead. The results range from around 16000 extra MMP control messages for the fastest moving hosts when n=20, to around 150 extra messages for the slowest moving single MH. Modified MMP generates around 20000 and 150 fewer messages than the default MMP for the two cases of MHs mentioned. Comparing the results for hops traversed by the control messages shows that they range from around 7000 extra hops in MMP case for the fastest moving MHs when n=20, to around 120 extra hops for the slowest moving single MH. Again, modified MMP performs better with around 21000 and 150 fewer hops than the default MMP for the two case mentioned. The same effect of a non-linear protocol overhead increase at the times which are multiples of the ECHO_INTERVAL observed in the case with the default MMP, can also be observed in Graph 4.10 and Graph 4.11. Since the interval was increased to 120.0 seconds this behaviour can be seen only once on the graphs causing a random increase in the protocol overhead for dwell times of 120.0 and 130.0 seconds (again results are more evident for the single MH case).

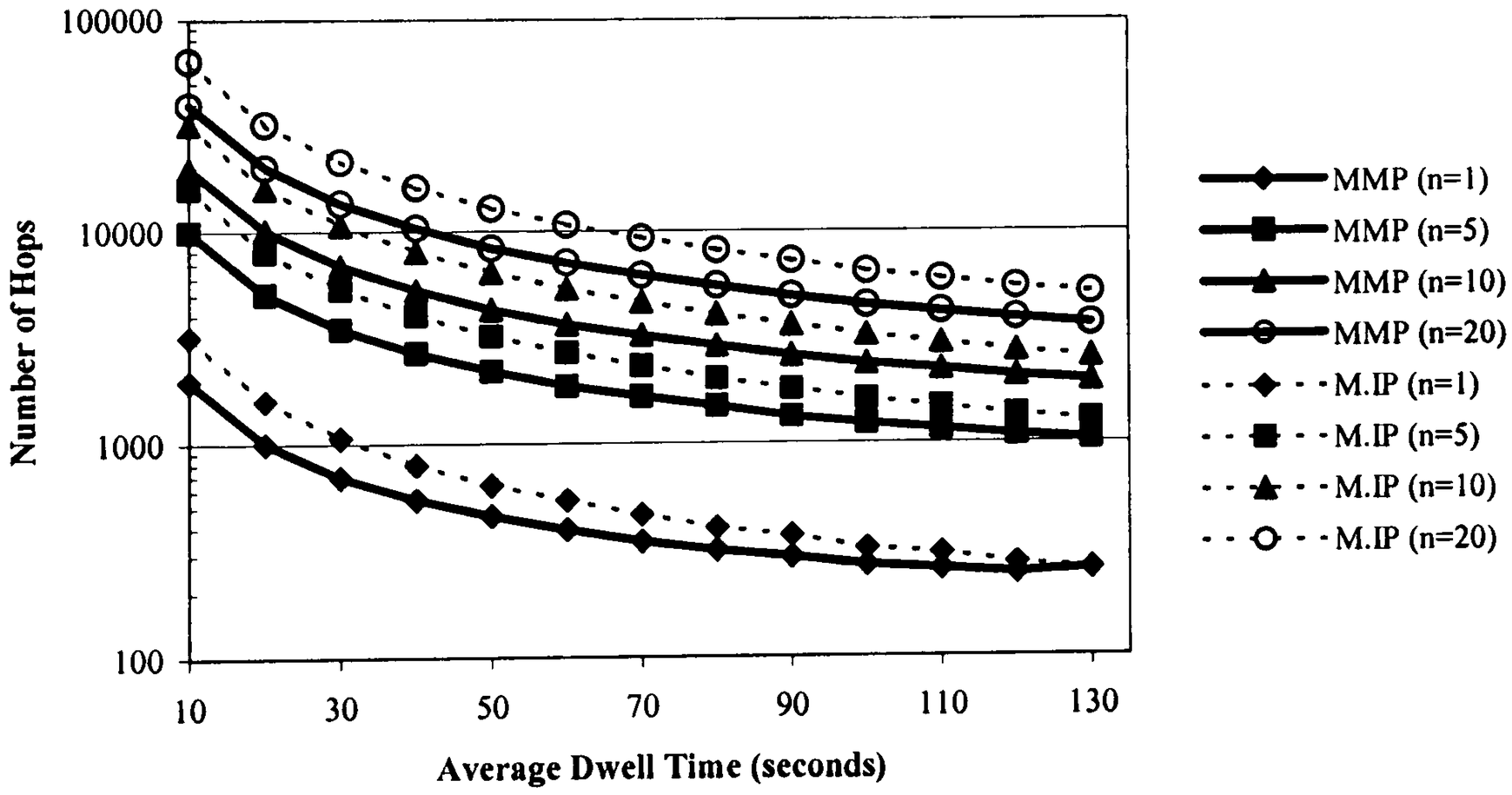


Graph 4.10. Control messages count: MMP versus Mobile IP (modified MMP and 1 Internet Hop), n is the number of MHs.

As indicated in the initial part of this chapter, Internet links (that is, outside the foreign network) are configured in such way that there is only one hop separating the Gateway/Core and the HA. This presents a very unlikely scenario since the possible separation between the visited and the home network of a MH is always random and can involve multiple hops.

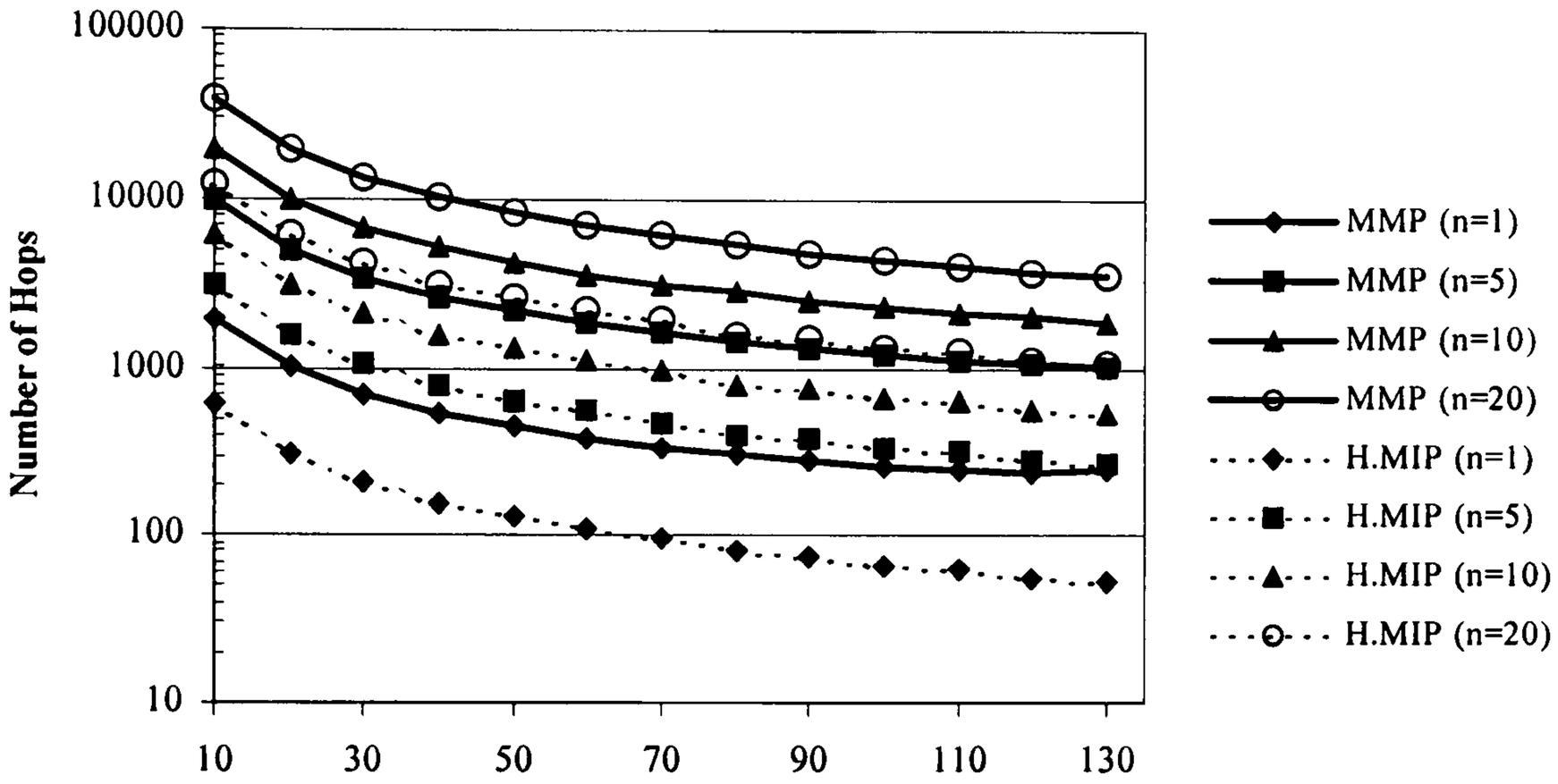


Graph 4.11. Hops count: MMP versus Mobile IP (modified MMP and 1 Internet Hop), n is the number of MHs.



Graph 4.12. Hops count: MMP versus Mobile IP (modified MMP and 5 Internet Hops), n is the number of MHs.

Graph 4.12 shows an extra scenario comparing modified MMP and Mobile IP for the number of hops traversed by the control messages. **The number of Internet hops for this case is set to 5**. The new value for Internet hops does not affect the protocol overhead in terms of the control messages generated. In this case, the effects of sending messages outside the foreign network are more obvious for Mobile IP since all Registration Requests and Registration Replies travel all the way to the HA for every handover while the MMP messages are heavily localised in the foreign network.



Graph 4.13. Hops count: MMP versus Hierarchical Mobile IP (modified MMP and 1 Internet Hop)

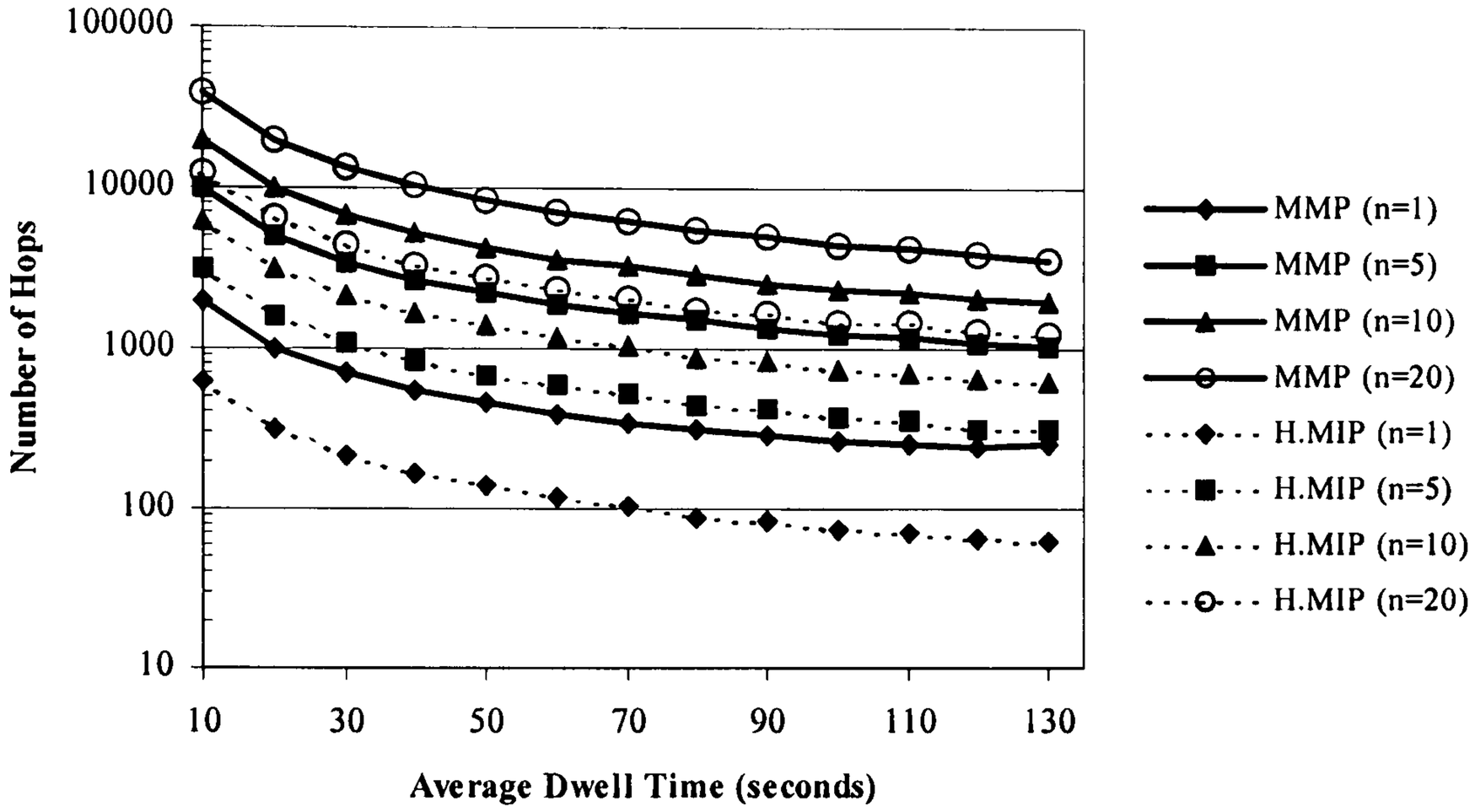
Hence, collectively, Mobile IP messages traverse more hops than MMP messages, although, as indicated in Graph 4.10 (which applies to this case since the number of control messages is not affected by the increase in the number of links) MMP generates more control messages in all population cases. In the case of the fastest moving MHs when n=20, Mobile IP messages traverse around 24000 more hops than MMP messages do, while for the slowest moving single MH the two protocols are almost equal. This is mainly due to the non-linear increase of MMP protocol overhead at the average dwell time of 120.0 seconds caused by the expiration of ECHO_INTERVAL. As Graph 4.12 shows, for the single MH, MMP performs better

for speeds which cause the host to stay in the cell for a shorter period than the ECHO_INTERVAL.

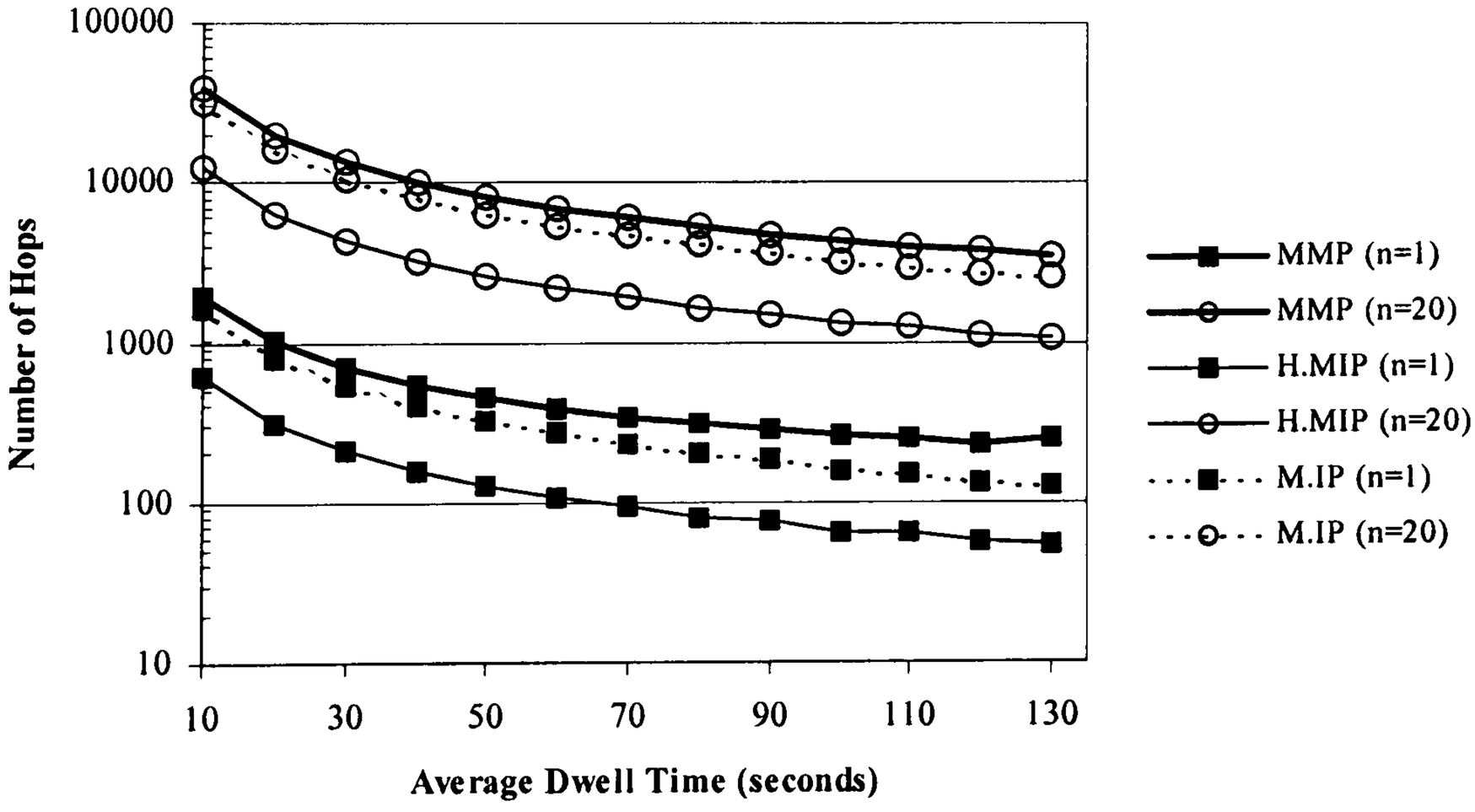
The protocol overhead of Hierarchical Mobile IP is obtained using the same simulation model and parameters as in the comparison of MMP and Mobile IP (Graph 4.8 to Graph 4.12). As explained in the initial part of the paragraph, transfer of Registration Requests and Replies is continuous: the FA on the route to the "cross-over" FA (or initially to the HA) replaces the source and destinations addresses of messages accordingly and hence does not create a new one for each step of the message transfer. Therefore, Hierarchical Mobile IP results are not different to the Mobile IP ones regarding the number of control messages generated during the lifetime of the connection since they use the same mechanisms for updating mobility databases[6]. The number of hops traversed by the control messages is the only comparison criteria and hence the results for Hierarchical Mobile IP (H. MIP in the legend) and MMP (modified MMP) are shown in Graph 4.13 and Graph 4.14 for the initial case of one internet hop and the case with five internet hops respectively. When the internet hops are set to one, the results range from around 25000 more hops for MMP messages for the fastest moving MHs when n=20, to around 180 more hops for MMP messages for the slowest moving single MH. When the Internet hops are set to five, results are almost identical with a slight, but inconsequential increase in the number of hops for Hierarchical Mobile IP. Due to local updates of "cross-over" FAs, Hierarchical Mobile IP performs even better than Mobile IP and it is not affected by the increase of the Internet hops since messages rarely travel all the way to the Home Agent. Finally, Graph 4.15 and Graph 4.16 show some population cases of MHs and the comparison of all three protocols tested (the results are extracted from cases already shown in earlier graphs of this chapter) regarding the number of hops that the

[6] As far as the behaviours of MHs with respect to the generation of control messages (especially in the wireless medium), Mobile IP and Hierarchical Mobile IP can be assumed to be the same. A similar statement was also made for MMP.

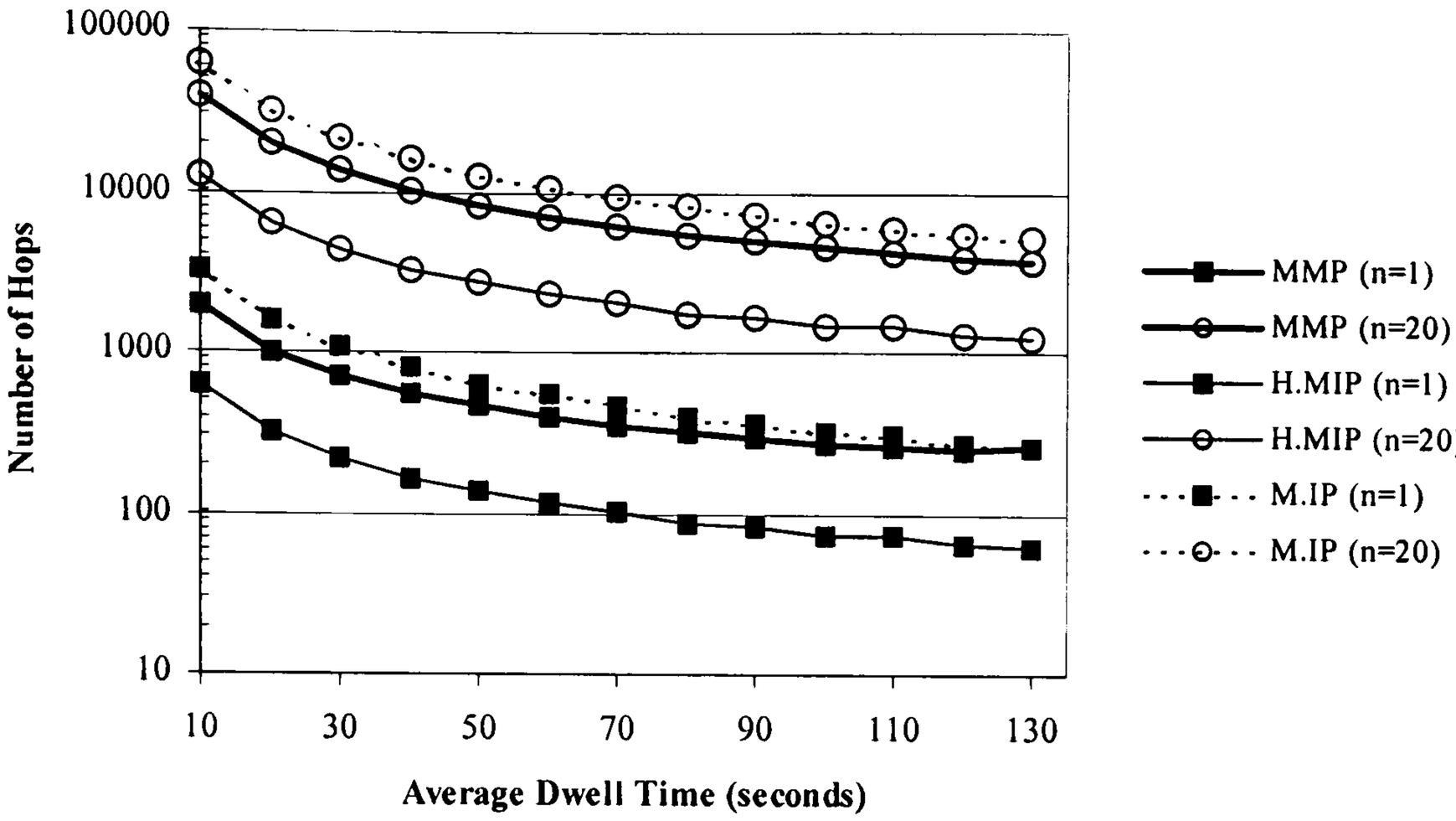
messages traverse for the two cases of different values of Internet hops. It can be clearly seen that, as far as the number of hops are concerned Hierarchical Mobile IP performs in the most efficient way. Further analysis and comparison between the protocols is presented in the next chapters. Finally, Graph 4.17 shows the effects of varying the values for Internet hops for all three protocols.



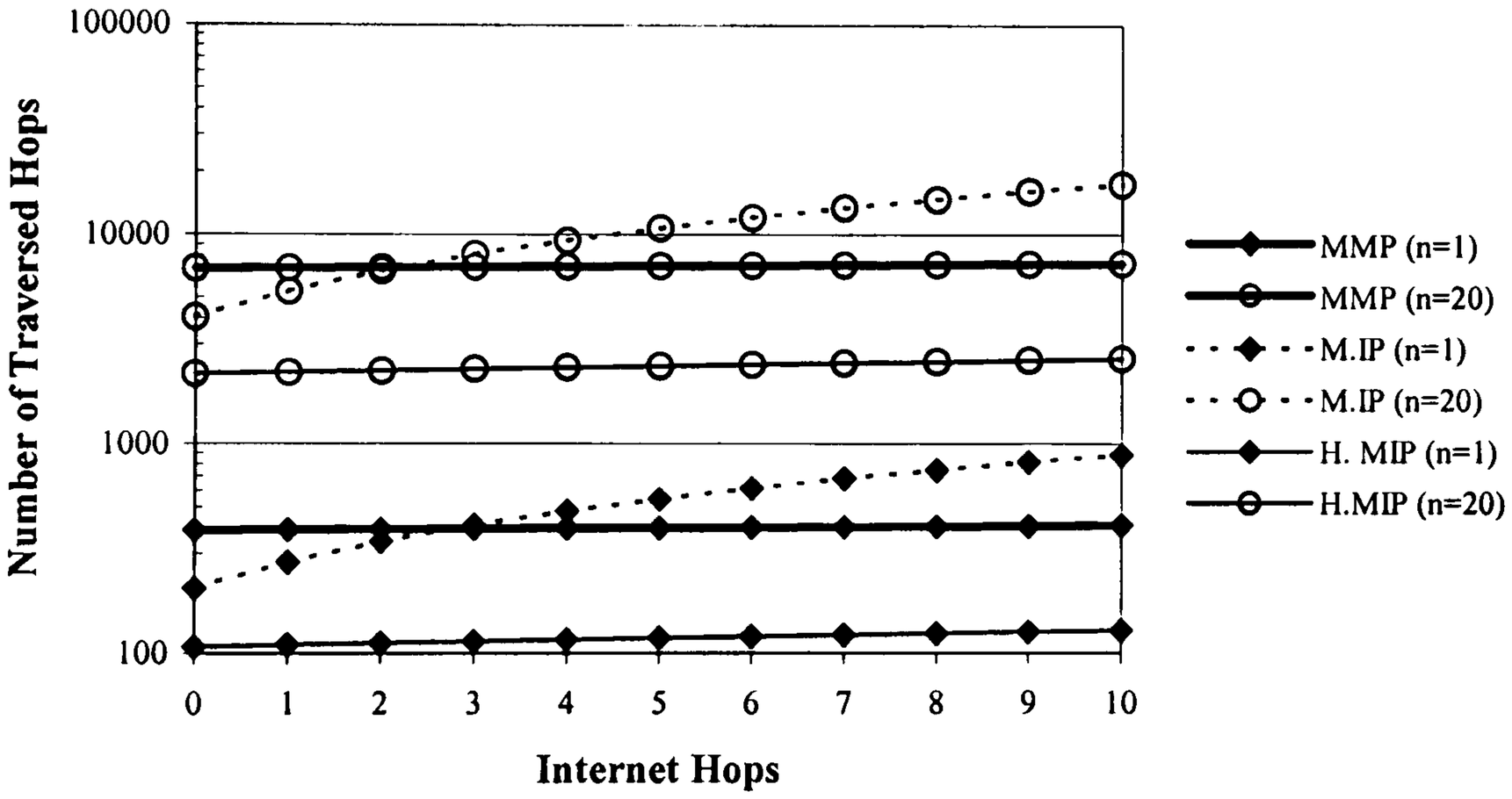
Graph 4.14. Hops Count: MMP versus Hierarchical Mobile IP (modified MMP and 5 Internet Hops)



Graph 4.15. Hops count: all three protocols (modified MMP and 1 Internet Hop)



Graph 4.16. Hops count: all three protocols (modified MMP and 5 Internet Hops)



Graph 4.17. Hops count: all three protocols. Effect of varying the number of Internet hops

4.4 Validation of Simulation Results

Behaviours of the simulated protocols shown in this chapter could be "observed" using the OPNET Modeller's Animation tool, which can produce animation files that capture and display operations of simulated protocols by showing network images and displaying movements of control messages and data packets. While the particular

OPNET animations can provide programmers with a reasonable level of confidence in the validity of conducted simulations, some further mathematical proofs can present a general validation method and further overview of protocol performances[7].

Results for the handover performance shown in section 4.2 show **packet losses** as consequences of the *handover latency,* which in turn is directly proportional to the *handover distance* for each handover case. From a perspective of protocol's performance, ***handover latency*** is the most relevant parameter since is indicates the time in which the mobility route can be diverted to the new point-of-attachment (i.e. BS). From an application's perspective **packet losses** are the main criteria in determining the efficiency and time needed for completing handovers. Naturally, packet losses and *handover latency* are mutually dependent. As indicated in Section 2.3.3 *handover latency* for IP mobility protocols is defined as the summation of the handover execution delay (movement detection) and the registration delay where the registration delay is the time needed to update to "cross over" router. If handover execution delay (movement detection delay) is defined as T_m and the registration delay is defined as T_r , the overall ***handover latency*** T_h in case of the simulated protocols becomes:

$$T_h = T_m + T_r \tag{4.1}$$

As discussed in the setup of simulations, gracious wireless handover is assumed to highlight the network layer effects of protocols' performances. Further considering that the movement detection procedures are identical for all simulated protocols (since they all utilise the Mobile IP movement detection procedure) effects of T_m are neglected thus making $T_m = 0$, which makes $T_h = T_r$.

[7] Mathematical validation and representation of IP mobility protocols is not extensively covered in the Internet research community mainly because of the recent emergence of the IP mobility protocols and the nature in which they are proposed and made public (i.e. IETF). Also, as pointed out at the beginning of this chapter mathematical models mostly serve the purpose of explaining the descriptive analysis of mobility protocols, which are often considered sufficient.

In order to explain the separation of ***handover latency*** and handover-incurred *packet losses* equation 4.2, suggested in [86], can be used to further present the concepts. As proposed in [86] packet losses defined as N_{loss} are related to the transmission rate of downlink traffic ω and handover loop time T_L defined as the transmission time from the "cross over" router to MH's old BS plus transmission time from MH's new BS to the "cross over" router:

$$N_{loss} = \omega \times T_L \tag{4.2}$$

The above equation assumes a particular situation installed in the simulations [86], which is to mark a particular packet by a "cross over" router. After receiving the marked packet the MH performs an immediate (gracious) handover to the new BS thus matching the equation's results with the number of lost packets to the closest value and making it directly proportional to the handover loop time T_L.

Going back to the definition of *handover latency* T_h in equation 4.1 the handover loop time in the conducted simulations, defined in equation 4.2, is equal to the summation of protocol *handover latency* and the cut-off delay in the old tree branch T_{cut} from the "cross over" router to the old BS. Thus,

$$T_L = T_h + T_{cut} \tag{4.3}$$

As shown in an example topology in Figure 4.6, in case of a handover with *handover distance* of one the handover loop time T_L is a summation of the Stage 1, which concerns the handover latency T_h , and the cut-off delay T_{cut} related to the Stage 2 of the handover. As already mentioned in the simulations shown in section 4.3, the discontinuity in the results shown for the tested protocol can sometimes occur because the traffic flow often does not entirely match the handover loop time. This was the main reason for the particular simulation strategy conducted in [86] where a packet

was marked by the "cross over" router inside the network, which was then used by MH to perform an immediate handover upon the packet's reception.

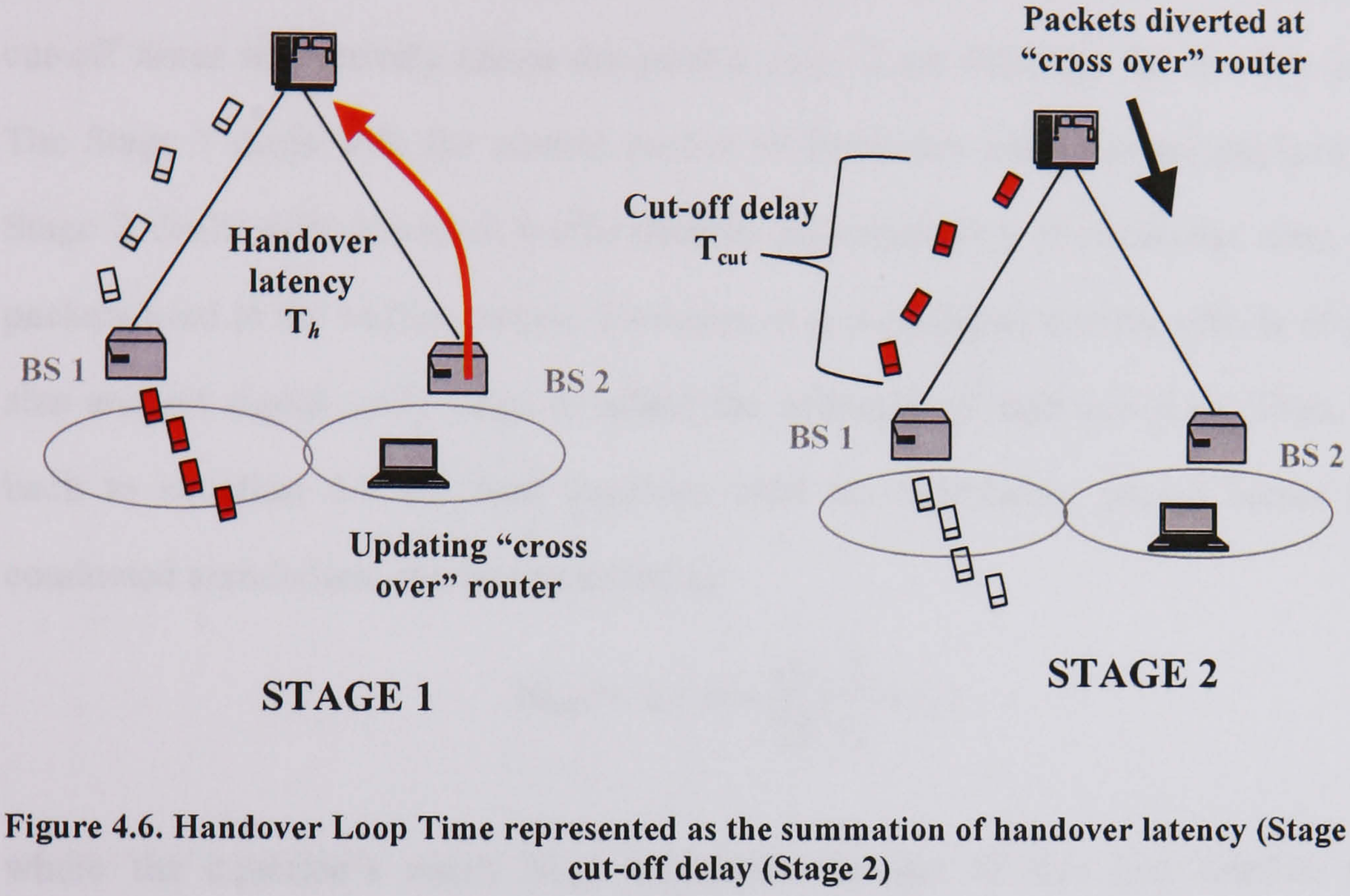


Figure 4.6. Handover Loop Time represented as the summation of handover latency (Stage 1) and cut-off delay (Stage 2)

Going back to the simulation setup applied in the conducted simulations, all links in the foreign network domain (micro mobility domain) in both cases of the *high* and *low bandwidth* networks contain identical characteristics including the transmission rates and delays. The implication of this is that the handover loop time in the conducted simulations can be approximated to, $\mathbf{T_L = 2T_h}$, since $T_{cut} = T_h$. Applying this statement into the calculation of the handover loop time for an arbitrary handover distance the following equation can be derived:

$$T_L = 2 \times \sum_{n=1}^{x} \left(\frac{S}{r_n} + t_n\right) \quad (4.4)$$

Where n is the particular hop of the *handover distance* starting with one from BS to Local Router in Figure 3.5 or LAN router in OPNET Network Editor of Figure 4.2, x is the *handover distance*, S is the packet size, r_n is the transmission rate of the hop and

t_n is the link delay set for the hop. Equation 4.4 reveals that the approximation $\mathbf{T_L=2T_h}$ is not considering the effects of packets sizes expressed in $\frac{S}{r_n}$ fraction and which are not the same for Stage 1 and Stage 2 related to the *handover latency* and cut-off times respectively (since the packet sizes S are different for the two stages). The Stage 1 deals with the control packet of fixed size and minimal payload while Stage 2 deals with downlink traffic packets corresponding to particular sizes of the packets used in the traffic models. However, it is considered that the effects of packet size are not significantly large to affect the accuracy of equation 4.4). Thus, going back to equation 4.1 the final equation used for calculating packet losses in the conducted simulations can be expressed as:

$$N_{loss} = 2 \times \omega \times \sum_{n=1}^{x} \left(\frac{S}{r_n} + t_n\right) \tag{4.5}$$

where the equation's result N_{loss}, represents number of bits lost relative to the particular rate of downlink packets flow. The equation only applies transmission and link delays used in the simulations since the processing delays are not simulated as already mentioned in the previous sections and propagation delays are considered negligible especially for point-to-point links.

Thus, if a particular point in Graph 4.1 is observed in case of MMP simulations for constant traffic models, *handover distance*, that is, $x = 3$, traffic rate $\omega = 1.024$ Mbits/s, packet size $S = 64$ bytes $= 512$ bits, r_1, , r_2 *and* $r_3 = 10$ Mbits/s, $t_1 = 0.5$ ms (bus link) and t_2 , $t_3 = 1.5$ ms. Where $n = 1$ is the first hop (bus link), $n = 2$ is the next hop upstream and $n = 3$ is the last upstream hop in the foreign network including the Gateway. Note: The packet size used is the size of data packets and bus links have an arbitrary delays of 0.5 ms (t_1) to represent the propagation and other delays on the bus link embedded in OPNET models of bus links. Hence:

$$N_{loss} = 2 \times 1024000 \times \left[\left(\frac{512}{10000000} + 0.0005\right)_1 + \left(\frac{512}{10000000} + 0.0015\right)_2\right.$$

$$+(\frac{512}{10000000}+0.0015)_3\,] = 2 \times 1024000 \times 0.0036536 = 7482.6 \text{ bits}$$

which, when divided with 512 bits for each packet gives 7482.6 / 512 = **14.614** lost packets (in bytes). The mathematical results closely match the simulations results of Graph 4.1 where for the result for MMP was **15** lost packets as shown on the graph for the particular point (last one in the curve for the *handover distance* of 3, No.1 in the Table 4.2). Some further results are shown in Table 4.2 for the *high-bandwidth* case.

No.	Protocol	Parameters	Equation 4.5/packet size (bits)	Simulation results (Graph 4.1)
1	MMP	x=3, ω = 1.024 Mbits/s, S=64 bytes,...(see above)	14.614	15
2	MMP	X = 3, ω = 512 kbits/s ...	7.3	6.5
3	MMP	X = 2, ω = 1.024 Mbits/s ...	8.4	6.5
4	MMP	X = 1, ω = 1.024 Mbits/s ...	2.2	0.5
5	Mobile IP[8]	x = 4, ω = 1.024 Mbits/s ,...,r_4 =30 Mbits/s, t_4 = 40ms	174.68	173.5
6	Mobile IP	x = 4, ω = 512 kbits/s,..., r_4 = 30 Mbits/s, t_4 = 40 ms	87.34	88

Table 4.2. Comparison of example simulation and mathematical results for the *high-bandwidth* network case

For the *low-bandwidth* network case new parameters are r_1, , r_2 *and* r_3 = 2.5 Mbits/s, t_2 , t_3 = 3 ms and r_4 = 7 Mbits/s. The rest of the parameters are the same as in the *high-bandwidth* case.

No.	Protocol	Parameters	Equation 4.5/packet size (bits)	Simulation results (Graph 4.3)
1	MMP	X=3, ω = 1.024 Mbits/s, ...	28.4576	28
2	MMP	X =3, ω = 512 kbits/s ...	14.2288	14
3	MMP	X =2, ω = 1.024 Mbits/s ...	15.6384	15
4	MMP	X =1, ω = 1.024 Mbits/s ...	2.8192	1
5	Mobile IP	X =4, ω = 1.024 Mbits/s ,...,r_4 =7 Mbits/s, t_4 = 40ms	188.75	186.6
6	Mobile IP	X =4, ω = 512 kbits/s,..., r_4 = 7 Mbits/s, t_4 = 40 ms	94.37	93.4

Table 4.3. Comparison of example simulation and mathematical results for *high-bandwidth* network case

[8] For Mobile IP results *handover distance* is constant x = 4 where the 4th (n=4) hop is the link between the Gateway and HA.

Regarding the exponential traffic models shown in Graph 4.2 and Graph 4.4 it was already mentioned that the average values of simulation results are identical for the constant traffic models thus maximum values are shown in the graphs. The match between the average values for the exponential traffic model and the validated values for the constant traffic model (and relative proximity of the maximum values shown in the graphs) is considered to provide enough confidence in the validity of the simulation results for the exponential traffic model.

Considering the simulation of different packets sizes shown in Graph 4.5, taking example packet size of 256 bytes the equation 4.5 gives 8.9 lost packet while the simulations result is 7 lost packets.

Based on the application of equation 4.5 for validating simulation results it can be observed that the mathematical and simulations result are more closely matched for cases of larger *handover distances.* One explanation of this occurrence is: as the delays in the calculations are accumulated for larger *handover distances* the accuracy increases between the simulations and mathematical validations because of the more frequent application of link delays (r_n) as the dominant delay factors which are significantly lager that the transmission delays. Also, a particular point of interest is the link delays used for bus links, which in OPNET simulations are set to the default OPNET values (fractions of *millisecond* per meter). It is considered that the applied value of t_l = 0.5 *ms* may be an overestimation of the actual bus link delays since the mathematical value for packet losses for small *handover distances* is larger that the one obtained in the simulations, suggesting an "inflated" handover loop time. However, this property is believed to serve the purpose of proving the mathematical model applied, as the simulation and mathematical results are not expected to be entirely identical due to the randomness of simulation results already pointed out in the non-integer values shown in the graphs. If an example is taken where the t_l was reduced to 0.1 *ms* this would replace the No.4 mathematical result shown in Table 4.2 with 0.6 closely matching the simulation result of 0.5 lost packets. As this actual proof

may appear to the reader as a sufficient indication that the $t_l = 0.1$ *ms* should be used in the mathematical validation this was deliberately avoided to point out the properties of the mathematical model.

Additionally, since the simulation results for Hierarchical Mobile IP are not shown due to their overlapping with some cases of *handover distance* for MMP the mathematical model applied for verifying the results for MMP and Mobile IP can be a sufficient proof that the simulations of Hierarchical Mobile IP are a true representation of protocol's behaviour in the test conditions.

Regarding the validation methods for the simulation results for the protocol overhead for the three tested protocols, the first point that needs to be considered is that the validation of the simulation results for the handover performance indicates that the protocols are simulated in an accurate way further signifying that the mechanisms of control messages and their processing in the network is correct. Observing the results for Mobile IP control messages count and hops they traverse the following equations can be applied respectively:

$$Number\ of\ control\ messages = \sum_{n}^{1,5,10,20} [(number\ of\ handovers + 1) \times 2] \qquad [4.6]$$

$$H_{(M.IP)} = Number\ of\ control\ messages \times (3 + Internet\ Hops) \qquad [4.7]$$

Where in equation 4.6, factor *number of handovers* +1 signifies the number of handovers in a simulation. Number of handovers is equal to the result of simulation time divided by dwell time (rounded up to the lowest integer value in case of a non-integer values and in case of the integer result, one is subtracted, since the last handover cannot be completed at the end of the simulation time) plus the initial login Registration Request/Reply exchange, which are multiplied by two for each Registration Request and Registration Reply for every handover. In equation 4.7 the number of control messages is multiplied by the number of hops they traverse this

being the path from BS to HA and it is equal to the summation of the three hops in the foreign network domain plus the value set for Internet Hops (one or five as set in the simulations). Finally, *n* is the number of MHs used in the simulation taking values 1, 5, 10 and 20. (Note: Mobile IP and Hierarchical Mobile IP simulations do not include the refreshments of entries, only handover related messaging)

Some exemplar results from equation 4.6 for the number of generated control messages are (for the simulation time of 2000 seconds):

(199+1)× 2 = 400 control messages (n=1, for 10 *ms* average dwell time, matches the result in Graph 4.8), 40 (n = 1, 100 *ms*, as in Graph 4.8), 336 (n = 5, 60 *ms* average dwell time distributed as 35.294, 40, 60, 120 and 200 ms, as in Graph 4.8), 508 (n=10, 80 *ms* average dwell time distributed twice as 46.15, 60, 80, 120 and 300 *ms*, as in Graph 4.8), 1344 (n=20, 60 *ms* average dwell time distributed four times as 35.294, 40, 60, 120 and 200 *ms*, as in Graph 4.8). Note: this proof is enough for all other graphs showing Mobile IP control messages generated since they are the same for all different simulated conditions, which mostly affect results for MMP and the number of hops traversed. Also as pointed in the description of Hierarchical Mobile IP simulations the protocol produces the same number of control messages, thus the same mathematical model applies.

Some exemplar results from equation 4.7 for the number of hops traversed by the control messages: 400×(3+1)=1600 (n=1, for 10 *ms* average dwell time, Internet Hops = 1, matches the result in Graph 4.9), 160 (n=1, 100 *ms*, Internet Hops = 1, as in Graph 4.9), 1344 (n=5, 60 *ms* average dwell time distributed as 35.294, 40, 60, 120 and 200 *ms*, Internet Hops = 1, as in Graph 4.9), 2032 (n = 10, 80 *ms* average dwell time distributed twice as 46.154, 60, 80, 120 and 300 *ms*, Internet Hops = 1, as in Graph 4.9), 5376 (n= 20, 60 *ms* average dwell time distributed four times as 35.294, 40, 60, 120 and 200 *ms*, Internet Hops = 1, as in Graph 4.9). Same results are included in other graphs with the same parameters. Further applying the same equation 4.7 but

with Internet hops = 5 renders 400×(3+5)=3200 (n =1, for 10 ms average dwell time, Internet Hops = 5, matching the result in Graph 4.12).

Regarding the validation method for MMP this introduces a complex challenge due to the use of protocol timers for transmissions of some control messages and the introduction of bus links in the simulations. As already indicated in the explanation of MMP simulations in section 4.3, bus links introduce some non-linear behaviours of protocol message transmissions due to the fact that a MH can hand over to adjacent BS, which is on the same bus link as the old BS. In this case the new BS sends a *MMP Instruct* message to the old BS to cut-off the old tree branch, which triggers a Quit Notification message sent upstream to the Local Router (LAN router in OPNET Network Image) being the uplink router for the bus link. The Local Router sets up a timer before deleting the upstream tree since the MH is still using the tree from the new BS. The new BS, upon receiving the Quit Notification sent by the old BS (since it is on the "broadcast" bus link), re-transmits the Join Request to make sure that the multicast tree is not pruned. In addition to depending on the value for MAX_RTX constant (default value set to 3) old BS sends MAX_RTX number of Quit Notifications, which in turn triggers the MAX_RTX number of Join Requests from the new BS. If a validation is attempted for the simplest case observed (meaning the least number of control messages generated) then a point in Graph 4.10 could be observed for n=1 and average dwell time of 130 seconds (resulting in the least number of handovers) for modified MMP where MAX_RTX = 1 (meaning only one transmission of Quit Notification from the old BS upon the reception of *MMP Instruct*) and ECHO_INTERVAL of 120 seconds. The process of estimating the number of control messages generated is as follows:

MH performs 2000/130 = 15.38 => 15 handovers plus the initial login. Thus:

16 *Registration Requests* are generated,

1 *Registration Reply* is returned by HA upon the login as the first *Registration Request* is proceeded to HA (others are received by Gateway with no Replies sent back),

16 *Join Requests* are sent for 15 handovers and initial login,

16 *Join Replies* are sent by "cross over" router for 15 handover and 1 login-triggered *Join Request,*

15 *MMP Instructs* are sent after 15 handovers to old BS and

15 *Quit Notifications* are sent in response to the reception of *MMP Instruct* (only one sent since MAX_RTX = 1)

There are 11 handovers where the new BS is on the same link as the old BS (assuming MH starts the session at the leftmost BS). These are between BS 8 and BS 9 in Figure 4.2 (or BS 1 and BS 2 of Figure 3.5), BS 8 and BS 10, BS 11 and BS 12, BS 12 and BS 13, BS 14 and BS 15, BS 15 and BS 16, BS 17 and BS 18, BS 18 and BS 19, BS 19 and BS 18, BS 18 and BS 17, BS 16 and BS 15.Thus there are:

11 *Join Requests* sent in response to the reception of *Quit Notifications* on the same bus link sent to LAN/Local Routers (no *Join Acks* sent in response to this event)

Since ECHO_INTERVAL is set to 120.0 seconds it can be assumed that for each wireless attachment to a BS, the BS sends an *Echo Request* to the upstream router, which triggers the *Echo Reply* as the acknowledgement. Thus, there are approximately (since average dwell time is 130 this is slightly higher that 120 which is the triggering interval – ECHO_INTERVAL for *Echo Requests*):

16 *Echo Requests* for the first hop BS to LAN/Local router

16 *Echo Replies* accordingly for the first hop…

Regarding the second hop LAN/Local router to Site Routers there are

16 *Echo Request* and

16 *Echo Replies*

Regarding the third hop Site Routers to the Gateway there are

16 *Echo Request* and

16 *Echo Replies*

Finally this gives 186 control messages matching the result in Graph 4.10 for n =1 and average dwell time of 130 seconds. Further validating the simulation results for MMP

control messages count the above result can be used as a reference. Taking the same case of Graph 4.10 where n =1 and for smaller average dwell times the reduction in generated control messages is less drastic for larger dwell times since all handover-related messages are similar (with slightly less number of handovers) and a small decrease in number of generated *Echo Requests* and *Echo Replies*. Note: this point for average dwell time of 130 is quite specific since more messages are generated than in the case slightly lower dwell time. This occurs because of the sudden triggering of Echo Requests after 120 seconds, which is too long for the other cases. Additionally, the results for n = 5, 10 and 20 are not multiples of the results for the case of n=1 simply because of the aggregation of *Echo Requests* and *Echo Replies* as specified in the description of MMP in Chapter 3. Similar behaviours can be extracted for default MMP with MAX_RTX = 3 and ECHO_INTERVAL = 60 seconds shown in other graphs.

If the same strategy is used for validating the results for number of hops traversed by the control messages the example of, n =1 and average dwell time of 130 seconds, can be used again since it generates the least number of control messages and its therefore easiest to validate for the hops the messages traverse. As proved before, 188 control messages are generated, which can be broken down in the following way to extract the number of hop control messages are traversing.

16 *Registration Requests* traverse: (1×4 hops+ 15×3 hops)= 49 hops (one goes to HA, others to Gateway),

1 *Registration Reply* traverses: 4 hops (from HA to BS) ,

16 *Join Requests* traverse: (2×3 hops + 11×1 hop + 3×2 hops)= 23 hops,

16 *Join Acks* traverse the same as Join Requests: 23 hops,

15 *MMP Instructs* traverse: (3×4 hops + 1×6 + 11×1 hop)= 29 hops,

15 *Quit Notifications* traverse: 15×1hops = 15 hops,

11 *Join Request* (*Quit Notification* triggered) traverse = 11×1 hop = 11 hops.

Since all Echo Requests and Echo Replies traverse a singe hop they can be accumulated. So they collectively traverse: 96 hops (see previous control messages count calculation).

This gives 250 as the final number of hops (for n=1, average dwell time 130 seconds, MAX_RTX = 1, ECHO_INTERVAL = 120.0 seconds) matching the simulation result in Graph 4.11. For other results for the number of hops traversed it can be observed that the increase in number of traversed hops is not linearly and inversely proportional to the reduction in cell dwell times. This was expected since the increase in control messages generated, thus the hop they traverse is handover related and these are mostly local messages, which traverse small number of hops. Additionally, as the average dwell time decreases transmissions of *Echo Requests* are reduced because of the smaller lifetime of the tree branches especially in the "lower" parts of the tree (i.e. closer to BSs). When the number of MHs is increased it is observed that the gradient of increase in hops traversed by the control messages is reduced as number of MHs is large (i.e. n=20) because of the more frequent overlapping of routing trees and thus *Echo Requests* and *Echo Replies* messages which for larger dwell times account for a large percentage of hops traversed.

4.5 Additional Performance Analysis

This section aims to extend the analysis of performances of simulated mobility protocols by applying both simulation results of section 4.2 and 4.3 and by expanding the mathematical models used for validation of the simulation results in section 4.4. The applied analysis is intended to provide some additional insight and more precisely explain the performance characteristics of the three mobility protocols obtained and described in the simulations. In addition, the applied deterministic analytical models developed, can assist in deducing some level of performance characteristics in additional network topologies such as the example topology considered in the next

section and used for further demonstration of the handover and protocol overhead performance analysis (additionally one more scenario of Hierarchical Mobile IP is introduced). The following sections are organised in the same way as the presented simulation results, firstly analysis is performed on the handover performances shown in the next section and then the protocol overhead is analysed in section 4.5.2.

4.5.1 Additional Handover Performance Analysis

One of the conclusions from the simulations of MMP, Hierarchical Mobile IP and Mobile IP is that MMP achieves the best handover performance (where Hierarchical Mobile IP is better than Mobile IP) by maximising the number of handovers with minimal possible *handover distance*. The simulation results reveal the magnitude of packet losses for individual handovers with their *handover distances*. In order to further understand the impact of different handover performances of the three tested mobility protocols this analysis observes the general impact of handovers on connectivity of MHs. Some of the simulation scenarios are used for determining the overall impact of handovers. This is then used for deriving a more generic parameter such as the average packets losses, which is then further exemplified in the additional network topology introduced in this section.

If a MH is performing the same sequence of events as used in simulations of handover performance, then the MH performs 11 handovers from the first BS to which it connects (BS_8 in Figure 4.2) to the rightmost BS (BS_19 in Figure 4.2) where it stops reception of packets. MH performs the following handovers with their associated *handover distances* for each mobility protocol tested:

- **Mobile IP**: A MH performs 11 handovers which have the same *handover distance* accumulating the distance between BSs and HA (these are 4 and 8 for the case where the Internet hops are set to 1 and 5 respectively).
- **Hierarchical Mobile IP**: A MH performs handovers with *handover distance* of 1 and 3. According to the setup of FAs used in simulation of Hierarchical Mobile IP

in section 4.2 MH performs 3 handovers with *handover distance* of 3. These are handovers between BS_10 and BS_11, BS_13 and BS_14, BS_16 and BS_17. The remaining 8 possible handovers have *handover distance* of 1 (see the end of section 4.2 for more detailed explanation).

- **MMP**: A MH performs handovers with *handover distances* of 1, 2 and 3. MH performs 1 handover with *handover distance* of 3. This is the handover between BS_13 and BS_14 shown in Figure 4.2. MH performs 2 handovers with *handover distance* of 2. In Figure 4.2 these are handovers between BS_10 and BS_11 and between BS_16 and BS_17. The remaining 8 possible handovers in the network have *handover distance* of 1.

The above can help in deducing that the benefit of MMP regarding the handover performance is in minimising *handover distances* for all performed handovers (due to the dynamic multicast routing tree adaptation to new BS from the "cross over" router) compared to the other two tested protocols. In addition, performance of MMP is fixed for the considered network topology unlike Mobile IP and Hierarchical Mobile IP where handover performance depends on the setup (i.e. locations) of FAs and HA (in the following analysis an additional scenario of Hierarchical Mobile IP is used to further highlight differences in handover performances and provide an additional example of this statement).

The simulation analysis of the handover performance of the three simulated protocols can be extended by observing handover packet losses for the whole duration of connectivity of a MH in the network. Hence, overall packet losses can be determined for the duration of a MH's connectivity in the network (this assumes active connectivity in the network where MH is engaged in receiving traffic as applied in all simulations in section 4.2). Hence, the simulation setup and results can be reused for calculating the overall packet losses for duration of a MH's connectivity in the foreign network. The overall packet loss experienced by a MH using the three simulated mobility protocols is defined as the summation of all packet losses for each handover

performed in the network. This can be examined for all different types of simulations performed for both *high* and *low-bandwidth* networks, different throughputs of the traffic destined to MHs and different traffic models. Hence if $N_{overall}$ is defined as the overall packet loss for duration of MHs connectivity in the network then for each protocol the results can be represented as follows:

$$N_{overall} = \sum_{m=1}^{n_{handovers}} N_{loss\,(h.d.),m} \tag{4.8}$$

Where $n_{handovers}$ is the total number of handovers performed by MH in the network and $N_{loss(h.d.),m}$ denotes the particular packet loss for the m^{th} handover performed with its corresponding *handover distance* (*h.d.*) applied in equation 4.5 in section 4.4. Applying equation 4.8 for the setup used in the simulations of handover performances in section 4.2 the following equations can be used for determining overall packet losses for simulated Mobile IP (M.IP), Hierarchical Mobile IP (H.MIP) and MMP:

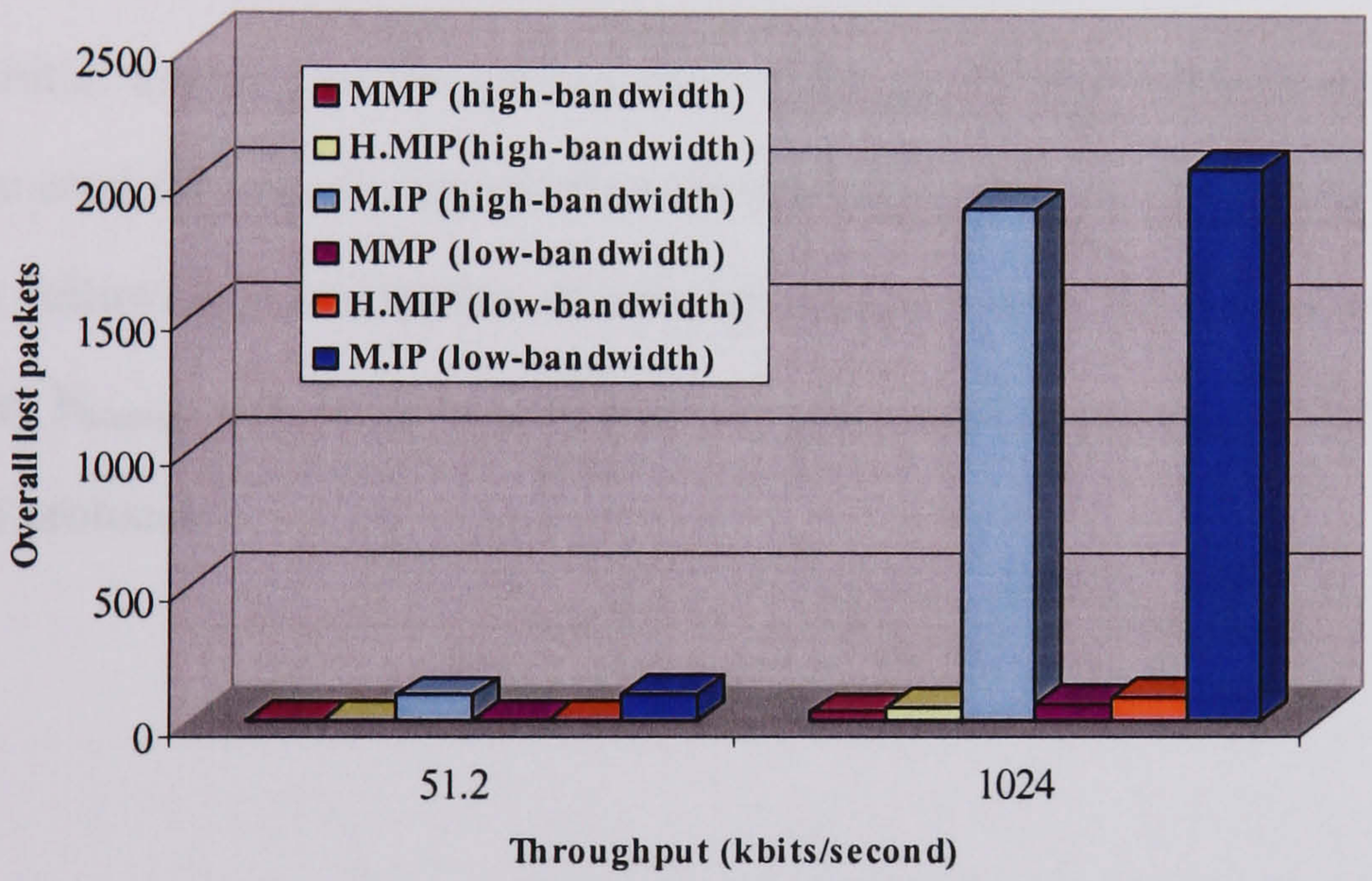
$$N_{overall}(M.IP) = 11 \times N_{loss(h.d.=4)} \tag{4.9}$$

$$N_{overall}(H.MIP) = (8 \times N_{loss(h.d.=1)}) + (3 \times N_{loss(h.d.=3)}) \tag{4.10}$$

$$N_{overall}(MMP) = (8 \times N_{loss(h.d.=1)}) + (2 \times N_{loss(h.d.=2)}) + (1 \times N_{loss(h.d.=3)}) \tag{4.11}$$

In case of Mobile IP (equation 4.9) all handovers have *handovers distance* of 4 following the setup used in the simulations where Internet hops are set to 1. Regarding equation 4.10 for Hierarchical Mobile IP and equation 4.11 for MMP, number of handovers correspond to the explanation at the beginning of this section. This principle of overall packet losses can be reused for any connectivity scenario of MHs defined in equation 4.8 considering number of handovers performed as can the average packet loss analysis shown in the latter parts of this section.

Graph 4.18 shows the overall packet losses for the simulated mobility protocols considering the cases of low and highest throughputs for MH traffic considered in the simulations ω=51.2 kbits/s and ω=1.024 Mbits/s for constant traffic model[9]. For the *high-bandwidth* network and ω=51.2 kbits/s values for MMP, Hierarchical Mobile IP and Mobile IP are 1, 1.5 and 93.5 overall lost packets respectively and for the same traffic rate the *low-bandwidth* network values for MMP, Hierarchical Mobile IP and Mobile IP are 3, 4.5 and 103.4 overall lost packets respectively. For the *high-bandwidth* network and ω= 1.024 Mbits/s values for MMP, Hierarchical Mobile IP and Mobile IP are 32, 49 and 1908.5 overall lost packets respectively and for the same traffic rate the *low-bandwidth* network values for MMP, Hierarchical Mobile IP and Mobile IP are 66, 92 and 2052.6 overall lost packets respectively.

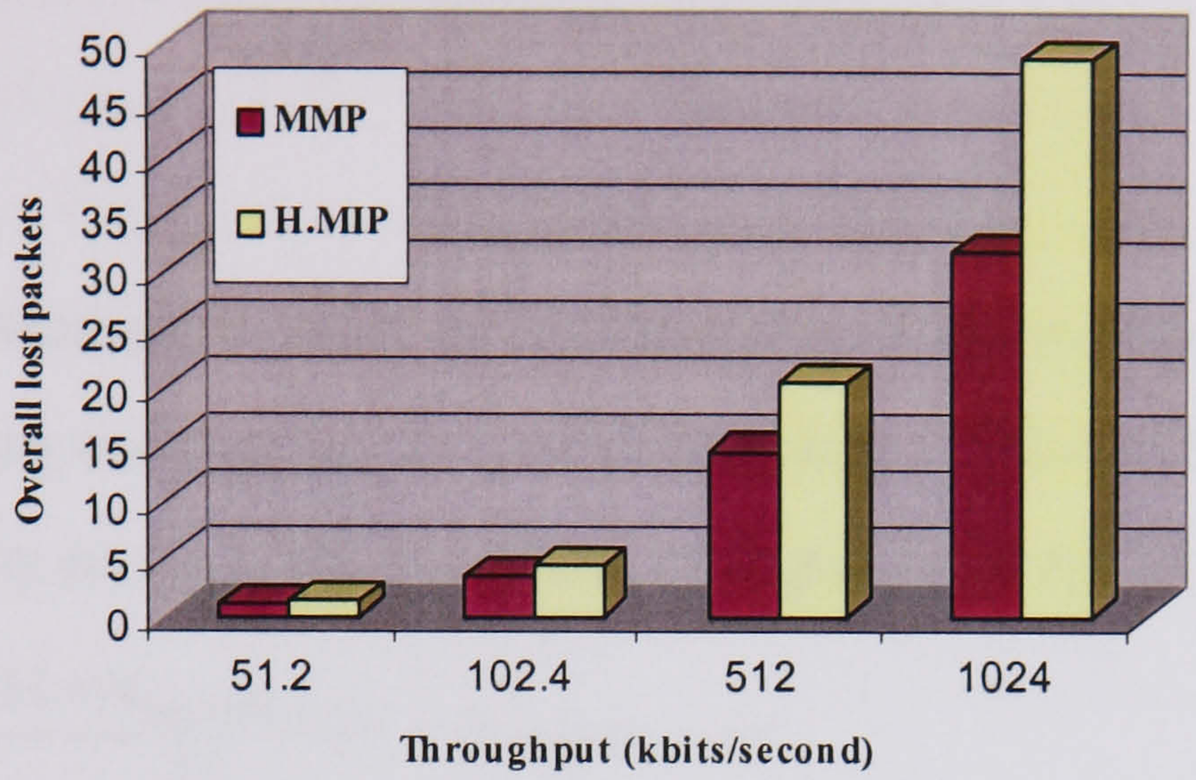


Graph 4.18. Overall packet loss for constant traffic model for the *high* and *low bandwidth* networks

Since Mobile IP experiences significantly larger overall packet losses for the examined cases, the differences between MMP and Hierarchical Mobile IP are less evident in Graph 4.18. Focusing on performances of MMP and Hierarchical Mobile IP

[9] Note: the lowest simulated throughput of 25.6 kbits/s was not considered since MMP and Hierarchical Mobile IP did not experience any packet losses in the simulations) for both *high-bandwidth* and *low-bandwidth* network defined in section 4.1.2.

Graph 4.19 shows more cases of traffic throughputs (from 51.2 kbits/s to 1024 kbits/s) for the constant traffic model and *high-bandwidth* network.



Graph 4.19. Overall packet loss for the constant traffic model for the *high-bandwidth* network (MMP and H.MIP only)

The initial analysis of the overall packet losses for MH connectivity in the network can be used for introducing another comparison criteria for the tested protocols, which is the following investigation of average packet losses for MHs using the protocols. Hence, $N_{average}$ can be generally defined representing average packet losses for each tested protocol:

$$N_{average} = \frac{N_{overall}}{n_{handovers}} \tag{4.12}$$

The accuracy of the results depends on the chosen method for calculating $N_{overall}$, where the previous analysis of overall packet losses and associated results is appropriate since MHs perform each possible handover in the network. Another way of defining $N_{overall}$ is to introduce $n_{handovers}$ in equation 4.8:

$$N_{overall} = n_{handovers} \times \sum_{m=1}^{n_{handovers}} \frac{N_{loss\,(h.d.),m}}{n_{handovers}} = n_{handovers} \times \sum_{l=1}^{h.d.} \frac{n_{handovers(h.d)} \times N_{loss\,(h.d.)}}{n_{handovers}} \tag{4.13}$$

Where $n_{handovers(h.d.)}$ is number of handovers for a particular *handover distance*. Putting this back to the equations 4.12 for $N_{average}$ obtains

$$N_{average} = \sum_{l=1}^{h.d.} \frac{n_{handovers(h.d)} \times N_{loss\,(h.d.)}}{n_{handovers}} \tag{4.14}$$

Hence, for each simulated mobility protocol the averages packet loss can be calculated using the following equations for the simulated topology (*handover distances* for each simulated mobility protocols are discussed at the beginning of this section):

$$N_{average}(M.IP) = \frac{11 \times N_{loss(MIP(h.d.=4))}}{11} = N_{loss(MIP(h.d.=4))} \tag{4.15}$$

$$N_{average}(H.MIP) = \frac{8 \times N_{loss(h.d.=1)}}{11} + \frac{3 \times N_{loss(h.d.=3)}}{11} \tag{4.16}$$

$$N_{average}(MMP) = \frac{8 \times N_{loss(h.d.=1)}}{11} + \frac{2 \times N_{loss(h.d.=2)}}{11} + \frac{1 \times N_{loss(h.d.=3)}}{11} \tag{4.17}$$

Graph 4.20. Average packet loss for constant traffic model for the *high-bandwidth* network

The above equations are used for plotting Graph 4.20 for each tested protocol in the *high-bandwidth* network. Mobile IP results are only shown for the traffic rate of ω=51.2 kbits/s since they are identical to the results shown in Graph 4.1 since Mobile IP handovers experience the same *handover distance* for every handover performed (also avoided for clarity of the other results for MMP and Hierarchical Mobile IP). Hierarchical Mobile IP experiences around 1.5454 more average packets losses than MMP for traffic rate of 1024 Mbits/s and for traffic rates of 512 kbits/s, 102.4 kbits/s and 51.2 kbit/s the differences are 0.545454, 0.090909 and 0.045455 respectively. The average packet losses can be used to deduce an estimate on the overall packet loss for arbitrary connectivity and session duration scenario with corresponding number of handovers. If this is checked against the values for overall packet losses in equation 4.12 in simulation scenarios for MMP and traffic rate of 1024 Mbits/s the value from Graph 4.20 is 2.9090909 which when multiplied with 11 handovers gives 32 lost packets matching the results in Graph 4.18. For MMP and traffic rate of 51.2 kbits/s the value from Graph 4.20 is 0.090909 which when multiplied with 11 handovers gives 1 lost packets matching the results in Graph 4.18. An additional example taken to point out the possible differences between MMP and Hierarchical Mobile IP is when a fast moving MH performs 40 handovers where for traffic rates of 1024 Mbits/s, 512 kbits/s, 102.4 kbits/s and 51.2 kbit/s the respective overall packet losses in MMP are: 116.3636, 52.7273, 12.7273 and 3.6363. For the same scenario Hierarchical Mobile IP gives the following overall packet losses: 178.1818, 74.5454, 16.36364 and 5.45456.

Regarding the average packets losses the theoretical ratios between the values for the mobility protocols considered should always be the same (assuming same packet sizes) since the average packet loss can also be represented by the handover latency, i.e. handover loop time defined in equation 4.4 and used for validation of simulation results in section 4.4. In addition, the whole analysis could be conducted taking the handover loop time instead of the packet losses (since packet losses are the product of

the traffic rate ω and handover loop time as shown in equation 4.5) assuming the same packet sizes as a factor in the equation 4.4 (see beginning of section 4.2). However, there are two main reasons why the shown ratios between the average packet losses for each traffic rate are not the same. Firstly, some of the values obtained from the simulations for smaller *handover distances* and lower traffic rates are zero as shown in section 4.2 thus equations 4.19 and 4.20 did not include all elements for the lower *handover distances*. Secondly, due to the random nature of simulation results (as discussed in section 4.2 simulation results are averages of 20 runs which also caused non-integer and values less than one for reasons such as the occurrence of "trapped" packets) there is always a small degree of offset which becomes an influencing factor when considered in the calculations of small values for packet losses. If the theoretical model is applied, the ratios would be the same but would not represent the simulated scenarios which are considered to provide more realistic insight into the performances of the protocols (in addition, the theoretical results are shown to provide accurate but not complete matching with the simulations results due to the above-mentioned reasons and issues discussed in section 4.4.)

The above analysis of the average packets losses can assist in extending the performance analysis of the simulated protocols for additional network topologies. One such foreign network topology is shown in Figure 4.7 with 24 BSs and the hierarchical topology converging at a single Gateway (the way this topology is formed is by merging two identical simulated topologies connected by the new fourth hop towards the new Gateway).

All three mobility protocols considered in this chapter can be applied to the **New Topology** shown in Figure 4.7. If a MH performs the same pattern of movement as applied in the simulations and also in the previous analysis of the three mobility protocols, then the MH executes 23 handovers (i.e. there are 23 handover situations in the network regarding the *handover distance*) from initial BS 1 to rightmost BS 24.

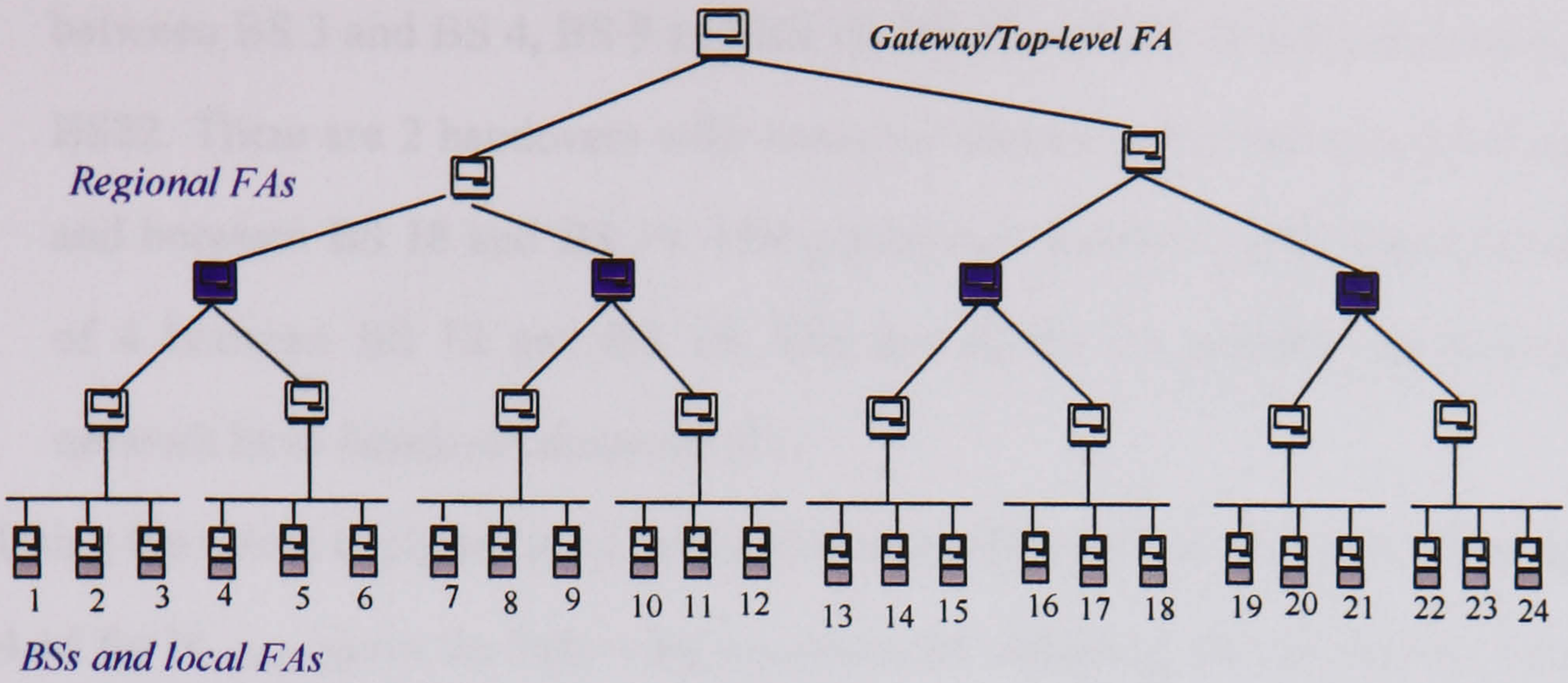


Figure 4.7. Additional Network Topology for the Foreign Network – New Topology

Hence, applying the same method used for the simulated topology, the following explains handover performances for each protocol:

- **Mobile IP**: A MH using Mobile IP performs handovers with *handover distance* accumulating the hop count from BS to the Gateway in the New Topology (4 hops) and number of Internet hops. Hence assuming 1 Internet hop as applied in the previous analysis for Mobile IP, MH performs 23 handovers with *handover distance* of 5.
- **Hierarchical Mobile IP**: If the same setup of FAs is applied to New Topology as used for the simulations and previous analysis of Hierarchical Mobile IP (see Figure 4.4 in section 4.1.2.1) then there are three levels of FAs in the foreign network. The top-level agent is placed in the Gateway, Regional FAs are placed two hops from the Gateways and Local FAs are placed in BSs (this is shown in Figure 4.7). Following this setup a MH using Hierarchical Mobile IP performs 3 handovers with *handover distance* of 4 between BS 6 and BS 7, BS 12 and BS 13 and between BS 18 and BS 19. The MH performs handovers with *handover distance* of two for the other 20 possible handovers in the network.

- **MMP**: A MH using MMP performs 4 handovers with *handover distance* of 2 between BS 3 and BS 4, BS 9 and BS 10, BS 15 and BS 16 and between BS21 and BS22. There are 2 handovers with *handover distance* of 3 between BS 6 and BS 7 and between BS 18 and BS 19. MH performs 1 handover with *handover distance* of 4 between BS 12 and BS 13. The rest of the 16 possible handovers in the network have *handover distance* of 1.

Using the above explanation of the handovers performed and applying this to equation 4.14 for $N_{average}$ gives the following equations for obtaining the average packet loss for each mobility protocol in New Topology:

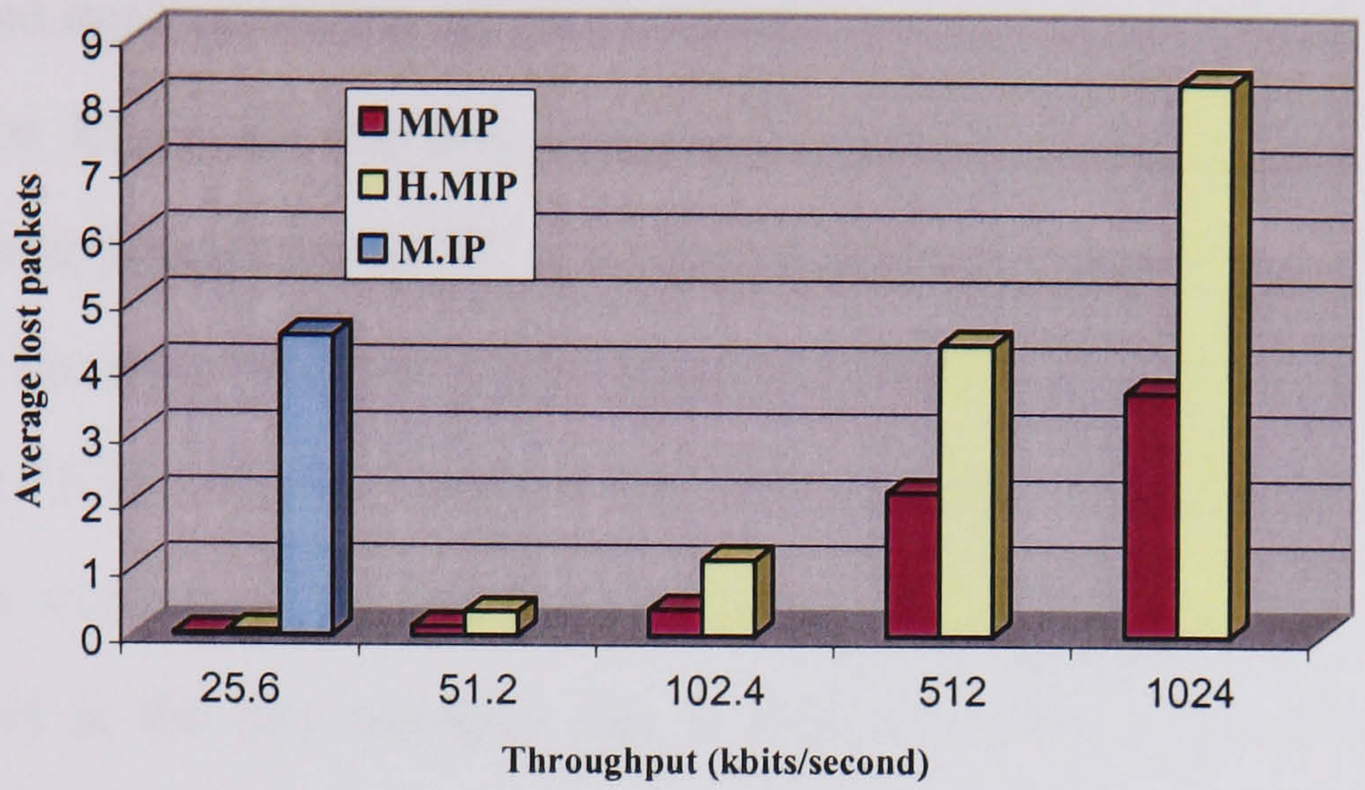
$$N_{average}(\text{M.IP\_new\_topology}) = \frac{23 \times N_{loss(MIP(h.d.=5))}}{23} = N_{loss(MIP(h.d.=5)} \tag{4.18}$$

$$N_{average}(\text{H.MIP\_new\_topology}) = \frac{20 \times N_{loss(MIP(h.d.=2))}}{23} + \frac{3 \times N_{loss(MIP(h.d.=4))}}{23} \tag{4.19}$$

$$N_{average}(\text{MMP\_new\_topology}) = \frac{16 \times N_{loss(MIP(h.d.=1))}}{23} + \frac{4 \times N_{loss(MIP(h.d.=2))}}{23} + \frac{2 \times N_{loss(MIP(h.d.=3))}}{23} + \frac{1 \times N_{loss(MIP(h.d.=4))}}{23} \tag{4.20}$$

The same parameters can be taken as in the simulated topology assuming that the new fourth hop to the Gateways has the same parameters as other point-to-point links in the network as done for the simulated topology. If *high-bandwidth* network case is assumed then for traffic throughput cases of ω = 25.6, 51.2, 102.4, 512, 1024 kbits/s packet losses for the new *handover distance* in the foreign network can be calculated using equation 4.5 where $N_{loss(h.d.=4)}$= 0.52048, 1.04096, 2.08192, 10.4096, 20.8192 packets lost respectively. At the same time, Mobile IP handover with *handover distance* of 5 (Internet hop set to 1) has $N_{loss(h.d.=5)}$ = 4.5222 lost packets for ω = 25.6 kbits/s. The average packet losses for the three mobility protocols are shown in Graph

4.21 (Mobile IP is only shown for ω = 25.6 kbits/s since the average is equal to the value given by equation 4.18 for *handover distance* of 5).

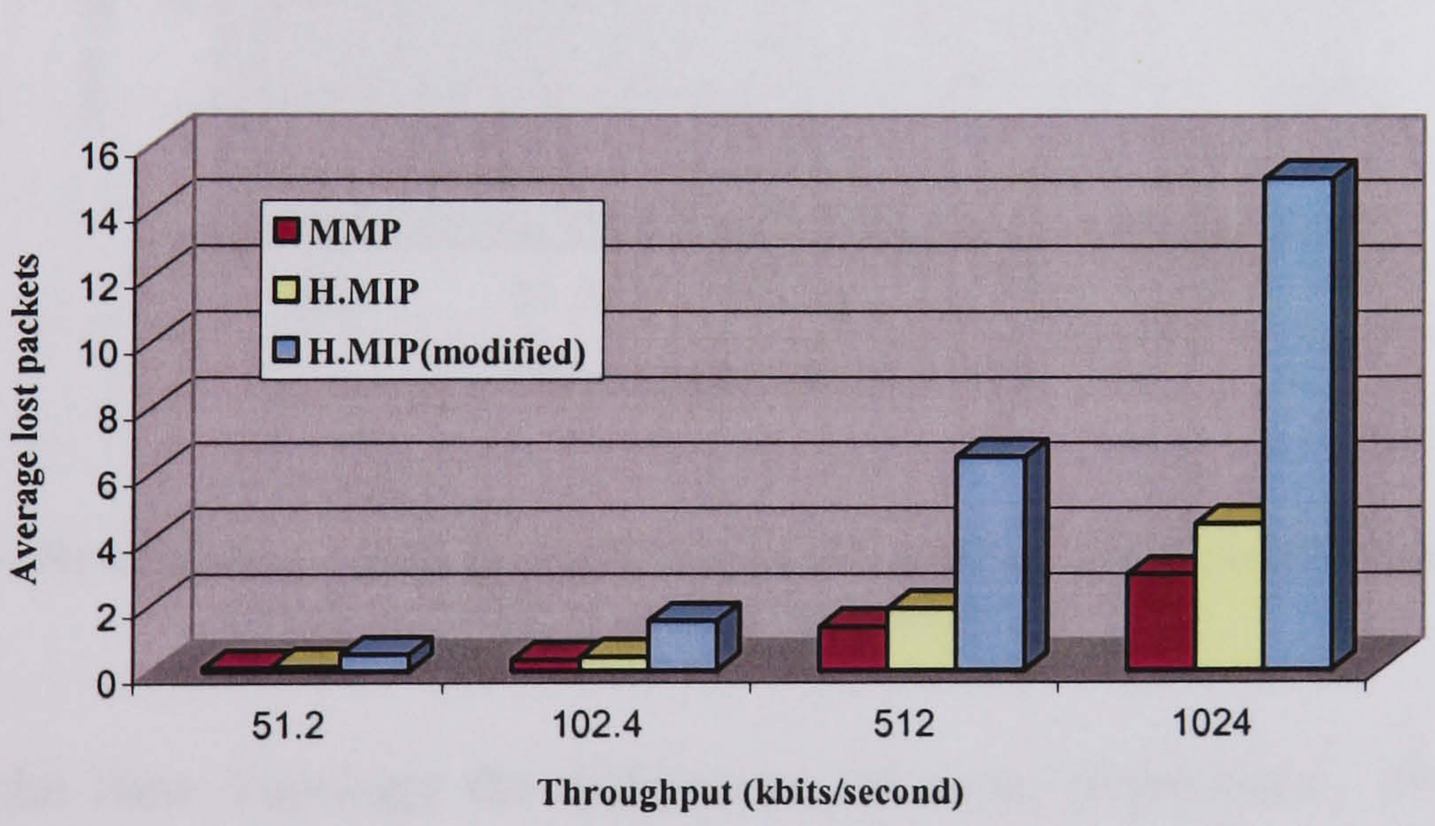


Graph 4.21. Average Packet losses for New Topology for constant traffic model and *high-bandwidth* network parameters

In case of the New Topology the differences between Hierarchical Mobile IP and MMP are increased compared to the simulated topology. For traffic rates of 1024 Mbits/s, 512 kbit/s, 102.4 kbits/s, 51.2 kbits/s and 25.6 kbit/s the respective differences in average packet losses are: 4.679931, 2.235201, 0.746235, 0.2209465 and 0.045259. If the average values are taken for calculating the overall packets loss according to the equation 4.12 for a MH performing 23 handovers (one full path in the New Topology) and traffic rate of 1024 Mbits/s MMP has 84.81 overall packet losses while Hierarchical Mobile IP has 192.457 overall packets losses (in the case number of handovers is increased to 40, MMP has 147.51 overall lost packets and Hierarchical Mobile IP has 334.71 overall lost packets).

The analytical tools used in the previous analysis can be extended to introduce another scenario of Hierarchical Mobile IP in addition to the one used in the simulations and previously shown calculation of packet losses. In this new setup for Hierarchical Mobile IP, only two levels of FAs are used this being the Gateway as the top level router and BSs as local FAs. Hence, this Hierarchical Mobile IP scenario excludes the

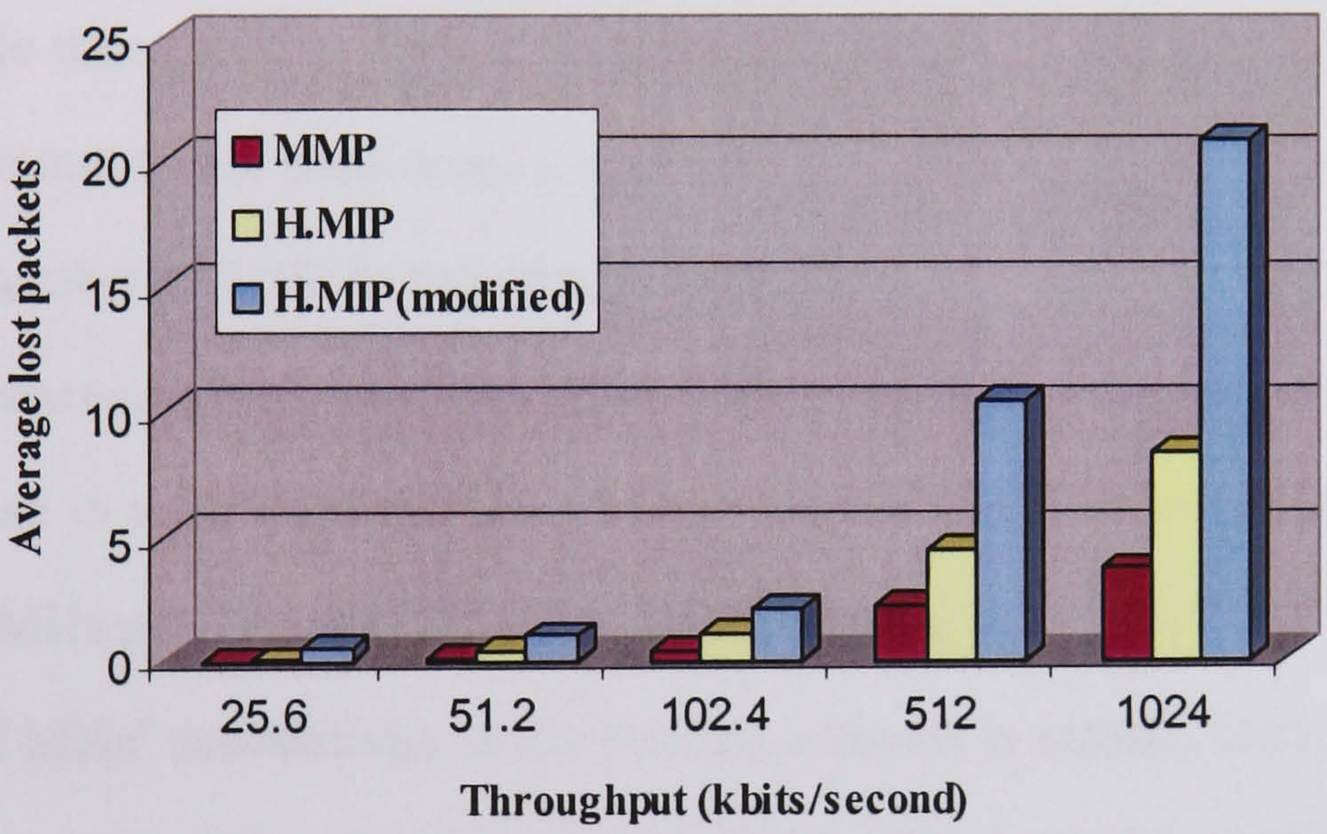
existence of regional FAs shown for both simulated topology and new topology. This presents a realistic scenario for Hierarchical Mobile IP and presence of a single routing "anchor" in the foreign network (beside the BSs) has already been considered as a design and implementation option for Hierarchical Mobile IP [33]. Calculation of average packet losses for this new scenario of Hierarchical Mobile IP, i.e. named **Modified Hierarchical Mobile IP**, is simplified because all handovers in the foreign network have the same *handover distance* corresponding to the hop count from BSs to the Top-level FA (i.e. the Gateway). In the topology used in the simulations, average packet losses correspond to the packet losses for *handover distance* of 3 (i.e. $N_{loss(h.d.=3)}$) and in the new topology this is incremented by 1 resulting in average packet losses equal to the losses for *handover distance* of 4 ($N_{loss(h.d.=4)}$). Comparison of Modified Hierarchical Mobile IP with the previously applied Hierarchical Mobile IP scenario and MMP are shown in Graph 4.22 for the topology used in the simulations and for the new topology the results are shown in Graph 4.23.



Graph 4.22. Average packet losses including modified Hierarchical Mobile IP (constant traffic model, *high-bandwidth* network) in simulated topology scenario

The new modified Hierarchical Mobile IP performs worse than MMP and the simulated-setup for Hierarchical Mobile IP. For the simulated topology shown in Graph 4.22 and the traffic rates for 1024 Mbits/s, 512 kbits/s, 102.4 kbits/s and 51.2 kbit/s, modified Hierarchical Mobile IP has 12.09, 5.18, 1.18 and 0.409 more average

packet losses than MMP respectively. When modified Hierarchical Mobile IP is compared to the simulated-setup of Hierarchical Mobile IP, the differences are (for the same order of traffic rates starting from 1024 Mbits/s): 10.545, 5.18, 1.18 and 0.3636 average lost packets. Finally, using the average values for the modified Hierarchical Mobile IP in simulated-topology, overall packet losses can be obtained using equation 4.12 for arbitrary number of handovers for a MH. In the simulated topology when MH performs 11 handovers modified Hierarchical Mobile IP gives 165 overall lost packets for the traffic rate of 1024 Mbits/s (for 40 handovers overall packet loss is 600 packets).



Graph 4.23. New Topology: Average packet losses including modified Hierarchical Mobile IP (constant traffic model, *high-bandwidth* network)

In case of the New Topology the differences are more emphasised. For the New Topology performance shown in Graph 4.23 and the traffic rates of 1024 Mbits/s, 512 kbits/s, 102.4 kbits/s, 51.2 kbit/s and 25.6 kbits/s, modified Hierarchical Mobile IP has 17.1314, 8.24, 1.687, 0.9087 and 0.49785 more average packet losses than MMP respectively. When modified Hierarchical Mobile IP is compared to the simulated-setup of Hierarchical Mobile IP the differences are (for the same order of traffic rates starting from 1024 Mbts/s): 12.45, 6.01, 0.94, 0.68779 and 0.45259 average lost

packets. Finally, using the average values for the modified Hierarchical Mobile IP in New Topology overall packet losses can be obtained for arbitrary number of handovers for a MH. In the New Topology when MH performs 11 handovers using equation 4.12 for the overall packet loss and traffic rate of 1024 Mbits/s, gives 229.01 overall lost packets (for 40 handovers overall packet loss is 832.8 packets).

4.5.2 Additional Protocol Overhead Analysis

The focus of this section is on the protocol overhead characteristics of MMP and its comparison with the other simulated mobility protocols by extending the simulation scenarios and observing the properties of the simulated protocols. One of the important goals of this analysis is to provide further analysis of the scalability of MMP alongside the other two simulated protocols: Mobile IP and Hierarchical Mobile IP. Several parameters are considered important for conducting the analysis:

- Increased number of MHs in the foreign network.
- As a consequence of the previous point density of MHs is increased in each cell. This increase is relative to the simulations scenarios where maximum number of MHs is 20 MHs per 12 cells in the simulated network.
- Duration of MHs' connectivity in the foreign network is varied. Thus, it is possible to determine protocol's operation for longer durations of time, which inevitably incurs more overall overhead.
- Different values for the protocol constants
- Some indication of the possible protocol overhead performances in a different network topology

The method used for validation of MMP protocol overhead results can be used as a starting point for developing mathematical models for scalability analysis of MMP. We can separate the generated messages in MMP into two categories considering whether they are related to individual behaviours of MHs or the aggregated collective property of the protocol:

1. **Movement-triggered messages**
2. **Soft-state messages**

Movement-triggered messages: These messages include Mobile IP-associated messages (*Registration Requests*) generated by MHs and their acknowledgments (*Registration Replies*) and *multicast tree-forming* messages defined in section 4.3 and described in chapter 3. The main characteristics of transmission of movement-triggered messages is that they are generated by each MH in the network according to protocol specifications and their overall overhead is the multiple of number of MHs in the network. The following derivation of movement-triggered messages for a single MH is done sequentially including and describing all steps in the message generations during handovers and login to the network. These can then be multiplied for number of MHs in the network. The focus is on the number of hops traversed (although the analysis always applies the number of control messages generated hence these can be also deduced) since these are relevant to the network topology and this is exemplified in the new topology introduced in the previous section and applied for this analysis at the end of the section. Hops traversed are network dependent and different for each protocol (section 4.3 describes that Mobile IP and Hierarchical Mobile IP generate the same number of messages).

During the **login** one *Registration Request* and one *Registrations Reply* are sent to HA by MH and acknowledged by HA respectively. This gives $2\times(d_{max\_foreign} + d_{internet})$ for the total number of hops the two messages collectively traverse in the foreign network and in the Internet. $d_{max\_foreign}$ is the maximum hop distance inside the foreign network between BSs and Gateways (in the simulated topology $d_{max\_foreign}= 3$) and $d_{internet}$ is the hops distance between the Gateway of the foreign network and HA of MHs (Internet hops as defined in section 4.3 where $d_{internet} = 1$ or 5). Additionally during the login, BSs generate and send a *Join Request* to the Gateway forming the transient multicast routing tree and the Gateway replies with a *Join Reply* forming the permanent multicast tree (see section 3.4.1). This gives

$$2 \times d_{max\_foreign}$$

for the number of hops traversed by these two messages.The total number of hops for the login phase of MH connectivity is

$$H_{login(MMP)} = 2 \times (d_{max\_foreign} + d_{internet}) + 2 \times d_{max\_foreign} =$$

$$= 4 \times d_{max\_foreign} + 2 \times d_{internet} \quad (4.21)$$

For values $d_{max\_foreign} = 3$ and $d_{internet} = 1$ equation 4.21 gives 14 hops traversed (this is used latter on for checking the validity of the model).

During **handovers** calculation of hops traversed is more complex as MH triggers CBT message generations in BS introducing several messages for supporting the transfer to the new BS. For each handover, MH sends one *Registration Request* to the Gateway (section 3.4.1 for explanation of the procedure). This gives

$$n_{handovers} \times d_{max\_foreign}$$

for the number of hops traversed for every MH for duration of its connectivity in the network. For each handover performed the previous step triggers *Join Request* generation by BS and transmission to the "cross over" router, which acknowledges it with a *Join Reply*. This gives

$$2 \times \sum_{m=1}^{n_{handovers}} d_m$$

for all hops traversed by CBT messages where d_m is the hop distance to the "cross over" router for each handover and is equal to the *handover distance* of the m^{th} handover performed in the system where $1 \le m \le n_{handovers}$. d_m takes values of 1, 2 and 3 since these are the possible *handover distances* in the simulated topology shown in Figure 4.2 and described in the simulation results in section 4.2. After every handover *MMP Instruct* is generated by new BS and sent to the old BS to "cut off" the old routing branch and stop the old BS from forwarding packets. Typically since *MMP Instruct* is sent to the old BS it would traverse $2 \times d_m$. hops for every handover performed since it would travel to the "cross over" router and back to the old BS (this

would be the case if all links in the simulated network are point-to-point links). Regarding the actual simulated topology the exception is during the handover with *handover distance* of 1 where bus links are used as done in the simulations in section 4.3 where *MMP Instruct* only traverses one hop since the old BS is on the same bus link and the message does not need to be sent back by the "cross over" router on that bus link (LAN router in Figure 4.2) since the old BS is able to receive the packet directly. This gives

$$\sum_{m=1}^{n_{handovers(h.d.>1)}} 2 \times d_m + n_{handovers(h.d.=1)}$$

for number of hops traversed by *MMP Instruct* messages where $n_{handovers(h.d.=1)}$ represents number of handovers for *handover distance* of 1 and $n_{handovers(h.d.>1)}$ represents number of handovers with *handover distance* larger than 1 (i.e. 2 and 3). For every reception of *MMP Instruct* a *Quit Notification* message is generated by BS and sent upstream to the next hop router. This is performed as a multiple of the constant MAX_RTX according to the MMP adaptations described in section 4.3. Finally this gives

$$n_{handovers} \times MAX\_RTX$$

for number of hops traversed by *Quit Notification* messages.

The above expression is MMP specific as explained in section 4.3. Due to the bus links used in the simulations, LAN routers on bus links in Figure 4.2 do not propagate the message immediately to the upstream routers but wait for a possible *Join Request* from the new BS and set a timer accordingly. This is done for allowing reception of a *Join Request,* which is sent by the new BS if MH performs handover to the new BS on the same bus link (this is analysed next). Since the intention was to show the effects of *MMP Instruct* message which triggers the *Quit Notifications* message even in the cases where MH performs handovers to different bus links the tree is assumed to time out during the timeout in LAN router which received the *Quit Notification* and it is not sent upstream (this was also applied in the simulation and for the validation of

simulation results in sections 4.4 and for proving the validity of the mathematical model below). Otherwise the above expression could be expressed as summation of MAX_RTX multiplied by the *handover distance* for each handover performed.

As explained in section 4.3 and above, according to CBT solutions for bus links, *Quit Notifications* received on bus links trigger transmission of *Join Requests* by MH's new BS to the upstream router on the same bus link without generation of *Join Acks* by the upstream router (in order to kept the established tree since MH is on the same bus link and uses the same upstream router). This gives

$$n_{handovers(same\_bus\_link)} \times MAX\_RTX$$

for the number of hops traversed by Join Acks specific to bus links where $n_{handovers(same\_bus\_link)}$ corresponds to the number of handovers where new and old BS are on the same bus link. As these are the only handovers with *handover distance* of 1 in network $n_{handovers(same\_bus\_link)} = n_{handovers(h.d.=1)}$. Hence the previous expression can be replaced with

$$n_{handovers(h.d.=1)} \times MAX\_RTX$$

Although the scenarios used in the latter part of the section deal with cases where every BS has at least one MH connected this would not trigger the transmission of *Join Request* from every BS since the message is multicast care-of-address specific and this would not be recognised by other BSs [18] (even if this is not the case and BS do not process the care-of-address other transmission could be suppressed and cancelled as in the case with soft-state messages described below).

Finally, total number of hops traversed during handovers becomes a summation of the above expressions for each message generated

$$H_{handover(MMP)} = (n_{handovers} \times d_{max\_foreign}) + (2 \times \sum_{m=1}^{n_{handovers}} d_m) + (\sum_{m=1}^{n_{handovers(h.d.>1)}} 2 \times d_m + n_{handovers(h.d.=1)})$$

$$+ (n_{handovers} \times MAX\_RTX) + (n_{handovers(h.d.=1)} \times MAX\_RTX) \quad (4.22)$$

Validity of equations 4.22 and 4.21 can be checked with the method applied in section 4.4 for "manual" calculation of the total hop count used for validating simulation results. Hence in the example value validated in section 4.4 $n_{handovers} = 15$ (dwell time of MH in the cell is 130 seconds), $d_{max\_foreign} = 3$ and $d_{internet} = 1$. Equation 4.21 obtains 14 hops traversed as previously shown while equation 4.22 can be calculated as follows:

$(n_{handovers} \times d_{max\_foreign}) = 45;$

$2 \times \sum_{m=1}^{n_{handovers}} d_m = 2 \times (3 + 3 \times 2 + 11) = 40$ (1 handover with d_m =3, 3 handovers with d_m =2, and 11 handovers with d_m =1);

$(\sum_{m=1}^{n_{handovers(h.d.>1)}} 2 \times d_m + n_{handovers(h.d.=1)}) = 6+3 \times 4+11= 29;$

$(n_{handovers} \times MAX\_RTX) = 15$ (for MAX_RTX=1);

$(n_{handovers(h.d.=1)} \times MAX\_RTX) = 11;$

Addition of the results of equation 4.21 (login bit) and 4.22 gives 14 + 45 + 40 + 29 + 15 +11 = 154 for the total number of hops traversed by the movement-triggered messages. The result used for validating simulation results in sections 4.4 is 250 for the overall number of hops traversed by MMP protocol including 96 hops caused by *Echo Request* and *Echo Reply* (these are not movement-triggered messages but related to soft-state as discussed in the latter parts of this section). Hence 154 + 96 = 250 proves the accuracy of equations 4.21 and 4.22.

Some further simplification of equation 4.22 can be applied to make it more generic and applicable to other network topologies (this is applied in the latter parts of this section for the example new topology). The summation $\sum_{m=1}^{n_{handovers}} d_m$ can be expressed as

$$d_{average} \times n_{handovers}$$

where $d_{average}$ is the average handover distance in a network. For the simulated topology having 11 possible handovers (one d_m =3, two d_m=2 and eight d_m=1)

$d_{average} = (1/11 \times 3 + 2/11 \times 2 + 8/11 \times 1) = 1.363636$

For the case when $n_{handovers} = 15$ considered previously

$$\sum_{m=1}^{n_{handovers}} d_m = 20$$

(see above) while

$d_{average} \times n_{handovers} = 20.4545.$

The accuracy increases for larger number of handovers and $n_{handovers} = 15$ is the simulated case with largest cell dwell time, i.e. least number of handovers. Hence when another example is taken $n_{handovers}=22$ (two full path across the network, an extreme simulated case with fastest movements includes 199 handovers for average dwell time of 10 seconds)

$$\sum_{m=1}^{n_{handovers}} d_m = (2\times 3 + 4\times 2 + 1\times 16) = 30$$

and when calculated using the new expression

$d_{average} \times n_{handovers} = 29.9999$

Some further simplifications can be done by taking the expression for calculating hops count of *MMP Instruct* messages

$$\left(\sum_{m=1}^{n_{handovers(h.d.>1)}} 2\times d_m + n_{handovers(h.d.=1)}\right) = 2\times \sum_{m=1}^{n_{handovers}} d_m - n_{handovers(h.d.=1)}$$

$$= 2\times (d_{average} \times n_{handovers}) - n_{handovers(h.d.=1)}$$

Using the new elements for equation 4.22 the new equation for calculating the hop count of handover messages becomes

$$H_{handover(MMP)} = (n_{handovers} \times d_{max\_foreign}) + 2\times (d_{average} \times n_{handovers}) + [2\times (d_{average} \times n_{handovers}) - n_{handovers(h.d.=1)}] + (n_{handovers} \times MAX\_RTX) + (n_{handovers(h.d.=1)} \times MAX\_RTX)$$

(4.23)

Finally, the total number of hops for the movement-triggered messages becomes summation of handover and login parts:

$$H_{movement\_triggered(MMP)} = H_{handover(MMP)} + H_{login(MMP)}$$

$$= (n_{handovers} \times d_{max\_foreign}) + 2 \times (d_{average} \times n_{handovers}) + [2 \times (d_{average} \times n_{handovers}) - n_{handovers(h.d.=1)}] + (n_{handovers} \times MAX\_RTX) + (n_{handovers(h.d.=1)} \times MAX\_RTX) + 4 \times d_{max\_foreign} + 2 \times d_{internet}$$

$$= 4 \times d_{max\_foreign} + 2 \times d_{internet} + n_{handovers} \times (d_{max\_foreign} + 4 \times d_{average} + MAX\_RTX) - n_{handovers(h.d.=1)} \times (1\text{-}MAX\_RTX)$$

(4.24)

Equation 4.24 can be checked in the same manner as done for equations 4.21 and 4.22 for the example value validated in section 4.4 where $n_{handovers} = 15$ (dwell time of MH in the cell was 130 seconds), $n_{handovers(h.d.=1)} = 11$, $d_{max\_foreign} = 3$ and $d_{internet} = 1$. Equation 4.24 gives the value of 155.818181, which gives 251.818181 when *Echo Request* and *Echo Reply* hops are added as explained in section 4.4 (Note: the slight increase is related to the introduced value for $d_{average}$ which is multiplied 4 times. Since this is the smallest number of handovers in the analysis below, the equation is assumed to produce more accurate results for other handover cases and is also useful for application to other topologies as exemplified for the new topology at the end of this section).

Regarding the value for $n_{handovers(h.d.=1)}$ this can be deduced manually by assuming the same movement pattern as in the simulations or can be expressed as the ratio for the simulated topology which is 8/11 (8 handovers with *handover distance* of 1 giving)

$$n_{handovers(h.d.=1)} = 8/11 \times n_{handovers}$$

hence this principle can be reused in any topology to avoid the manual insertion of the value for $n_{handovers(h.d.=1)}$ assuming the similar movement pattern as in the simulations for best accuracy.

Finally, the total number of hops for the movement-triggered messages in the network for arbitrary number of MHs in the network is given by the following equation:

$$H_{movement\_triggered(MMP)} = MH\_count \times [4 \times d_{max\_foreign} + 2 \times d_{internet} + n_{handovers} \times (d_{max\_foreign} + 4 \times d_{average} + MAX\_RTX) - n_{handovers(h.d.=1)} \times (1\text{-}MAX\_RTX)] \quad (4.25)$$

where *MH_count* is the number of MHs in the network using MMP.

Soft state messages: These include *multicast tree maintenance* messages defined in section 4.3 ("keepalive" messages) which are *Echo Requests* and their acknowledgments *Echo Replies* exchanged between the adjacent downstream-to-upstream routers. These messages are generated according to CBT multicast routing specification and are aggregated for all MHs using the particular routing branch (since *Echo Requests* do not carry multicast address information [18]). This constitutes the major difference between the soft state messages and movement-triggered messages as the former are aggregated for all MHs using a particular routing tree branch, i.e. their transmission is dependent on whether at least one MH uses the routing branch. At the same time, these messages are not generated by MHs but by the routers in the routing branch (i.e. from BS or any router apart from the Gateway to the next hop upstream neighbour). In order to derive the equation for calculating the number of hops traversed by the "keepalive" messages in the simulated topology, the first step is to determine number of hops (links) between routers in the network assuming all point-to-point links in the simulated topology (i.e. the first step is to assume that LAN routers are connected to each of the three BSs using a point-to-point link instead of the bus link). This is given by

$$\sum_{level=2}^{n} Node_{level}$$

where n is the total number of *levels* in the routing hierarchy. The hierarchy of the Gateway is *level* 1 down to BSs, which are *level 4* in the network used in the simulation of MMP. $Node_{level}$ represents number of *Nodes* (routers or BSs) in each *level* in the topology. Figure 4.8 represents the theoretical representation of the simulated network with bus links with the indication of *levels* and *Nodes* in each *level*.

In order to derive the modes for the bus links as they are used for *level 4* (i.e. BS) connections in the simulated topology, the first step is to observe the method applied in simulations and described in sections 3.4.3.1 as specified in CBT [18]. In bus links the first transmission of an *Echo Request* suppresses the scheduled transmissions in other BSs on the link. Hence the bus link can be assumed as a single link in the calculation of soft-state messages since only one pair of *Echo Request* and *Echo Reply* is exchanged regardless of the number of BS on the bus links (otherwise the above point-to-point model could be assumed and applied). Hence the final expression for the hops traversed in the simulated topology regarding the transmission of soft-state messages is given by

$$[\sum_{level=2}^{n} Node_{level}]- [Node_{(n-1)} \times (\frac{Node_n}{Node_{(n-1)}} -1)]$$

where $\frac{Node_n}{Node_{(n-1)}}$ represents number of possible point-to-point links between each the upstream router on bus links and BSs connected to it (i.e LAN router in Figure 4.2 with *level* = *n-1*). Since the expression started from the point-to-point links and in order to derive it for bus links used in the simulated topology the redundant point-to-point links are subtracted. These are $(\frac{Node_n}{Node_{(n-1)}} -1)$ and this is done for each bus link in the network hence it is multiplied by $Node_{(n-1)}$. For the topology shown in Figure 4.8 number of effective hops relevant to the soft-state calculation is given by the following expression, which can be manually checked in Figure 4.8 (note: $n = 4$, $Node_1 = 1$, $Node_2 = 2$, $Node_3 = 4$, $Node_4 = 12$)

$$[\sum_{level=2}^{n} Node_{level}]- [Node_{(n-1)} \times (\frac{Node_n}{Node_{(n-1)}} -1)] = 18 - [4 \times (3-1)] = 10$$

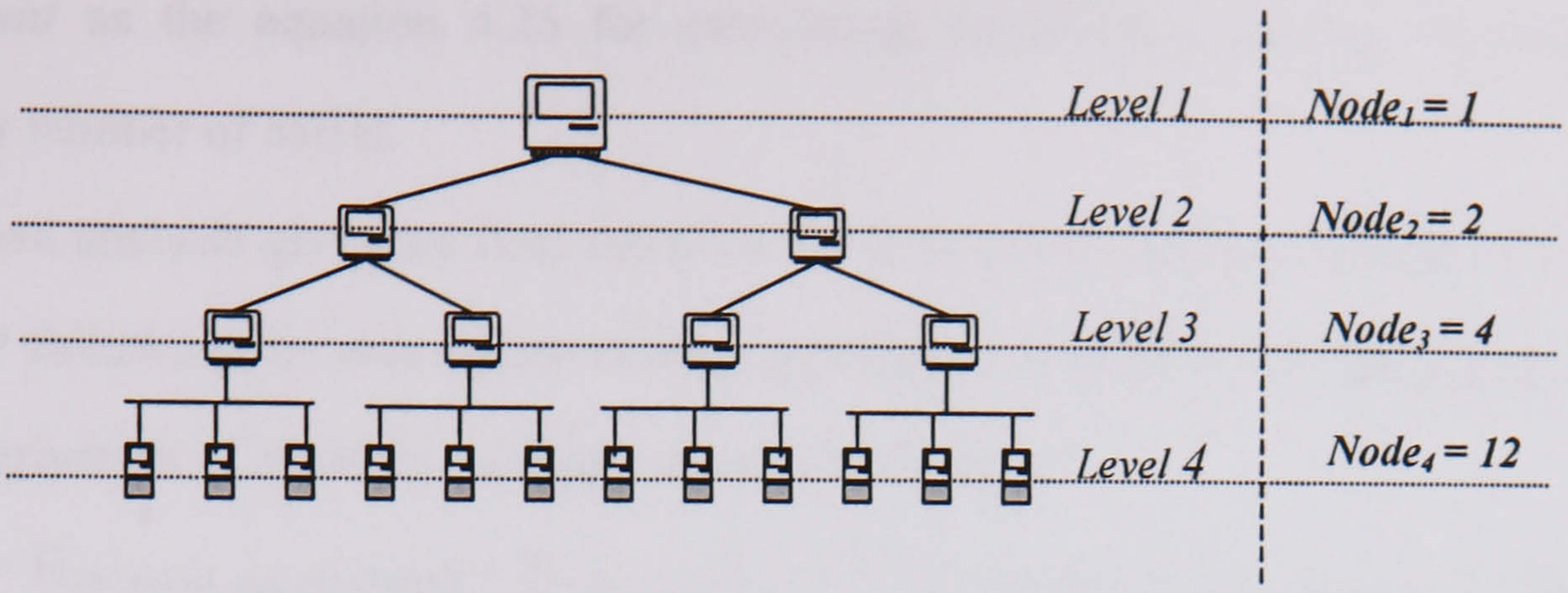


Figure 4.8. Topology parameters for calculation of network hops for the simulated topology

As described in section 3.4.3.1 and CBT specifications [18], *Echo Request* is sent to the next hop upstream where the upstream router acknowledges it with an *Echo Reply* thus the above number of effective hops is multiplied by 2. Finally assuming that the population of MHs in the network is such that there is always at least one MH presented in every cell (i.e. connected to every BS) this means that the whole network is maintaining the multicast routing tree. The final number of "soft state" messages is given by

$$H_{soft\text{-}state(MMP)} = 2\times\{[\sum_{level=2}^{n} Node_{level}\,]- [Node_{(n-1)}\times(\frac{Node_n}{Node_{(n-1)}}-1)]\times T_{overall} \times 1/T_{refresh\_interval}\} \quad (4.26)$$

Where $T_{overall}$ is the overall time for connectivity of MHs in the network also called overall dwell time in the simulations in section 4.3 (or any interval of interest and it equals to $n_{handovers} \times T_{averege\_dwell}$ where $T_{averege\_dwell}$ is the average dwell time in the cell as used in the simulations). $1/T_{refresh\_interval}$ is the frequency of refreshments, which corresponds to the ECHO_INTERVAL CBT constant [18] as used in simulations of MMP protocol overhead in section 4.3. *Echo Reply* is generated by reception of *Echo Request* as explained in section 3.4.3.1 and sections 4.3 and 4.4 (Note: due to the properties described for soft-state messages equation 4.26 is not multiplied by

MH_count as the equation 4.25 for calculating movement-triggered message for arbitrary number of MHs).

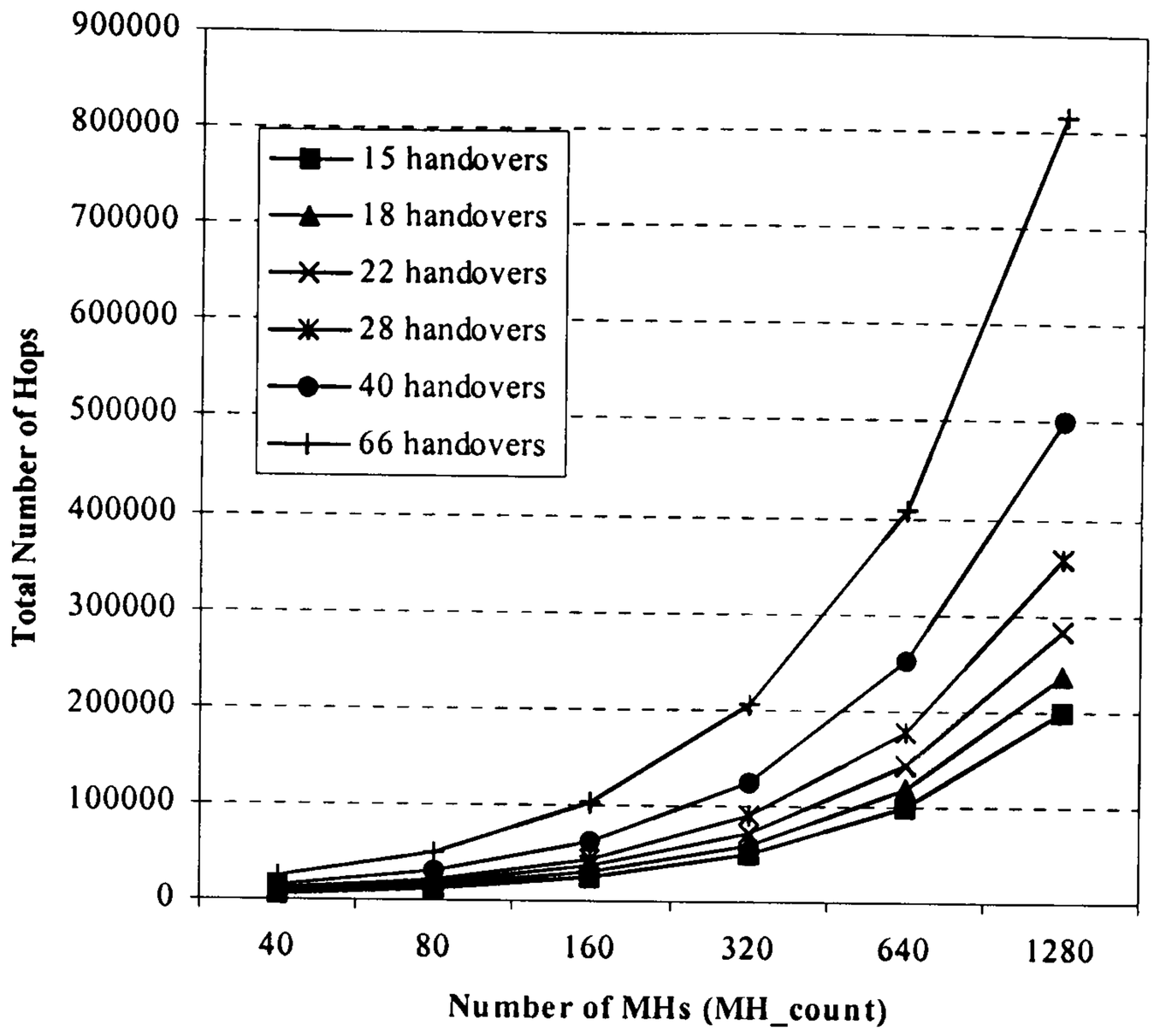
The above analysis gives the final equation for calculating number of hops generated in MMP following the assumption of high-population of MHs in the network applied in the derivation of equation 4.26 for soft-state messages:

$$H_{(MMP)} = H_{movement\_triggered(MMP)} + H_{soft\text{-}state\ (MMP)} = MH\_count \times [4 \times d_{max\_foreign} + 2 \times d_{internet} + n_{handovers} \times (d_{max\_foreign} + 4 \times d_{average} + MAX\_RTX) - n_{handovers(h.d.=1)} \times (1\text{-}MAX\_RTX)] + 2 \times \{[\sum_{level=2}^{n} Node_{level}] - [Node_{(n-1)} \times (\frac{Node_n}{Node_{(n-1)}} - 1)] \times T_{overall} \times 1/T_{refresh\_interval}\} \quad (4.27)$$

In the following the focus is initially placed on protocol overhead performance of MMP for increased population of MHs and on the implications of using multicast routing for solving mobility of MHs. The analysis is then extended to introduce comparisons with Mobile IP and Hierarchical Mobile IP. In the topology used in the simulations there is a maximum of 20 MHs distributed in 11 BSs in the network. The population cases used in the further performance analysis shown in this section are *MH_Count* = 40, 80, 160, 320, 640 and 1280 MHs in the network ranging from the average individual cell population of 3.6 to 116.36 MHs per cell.

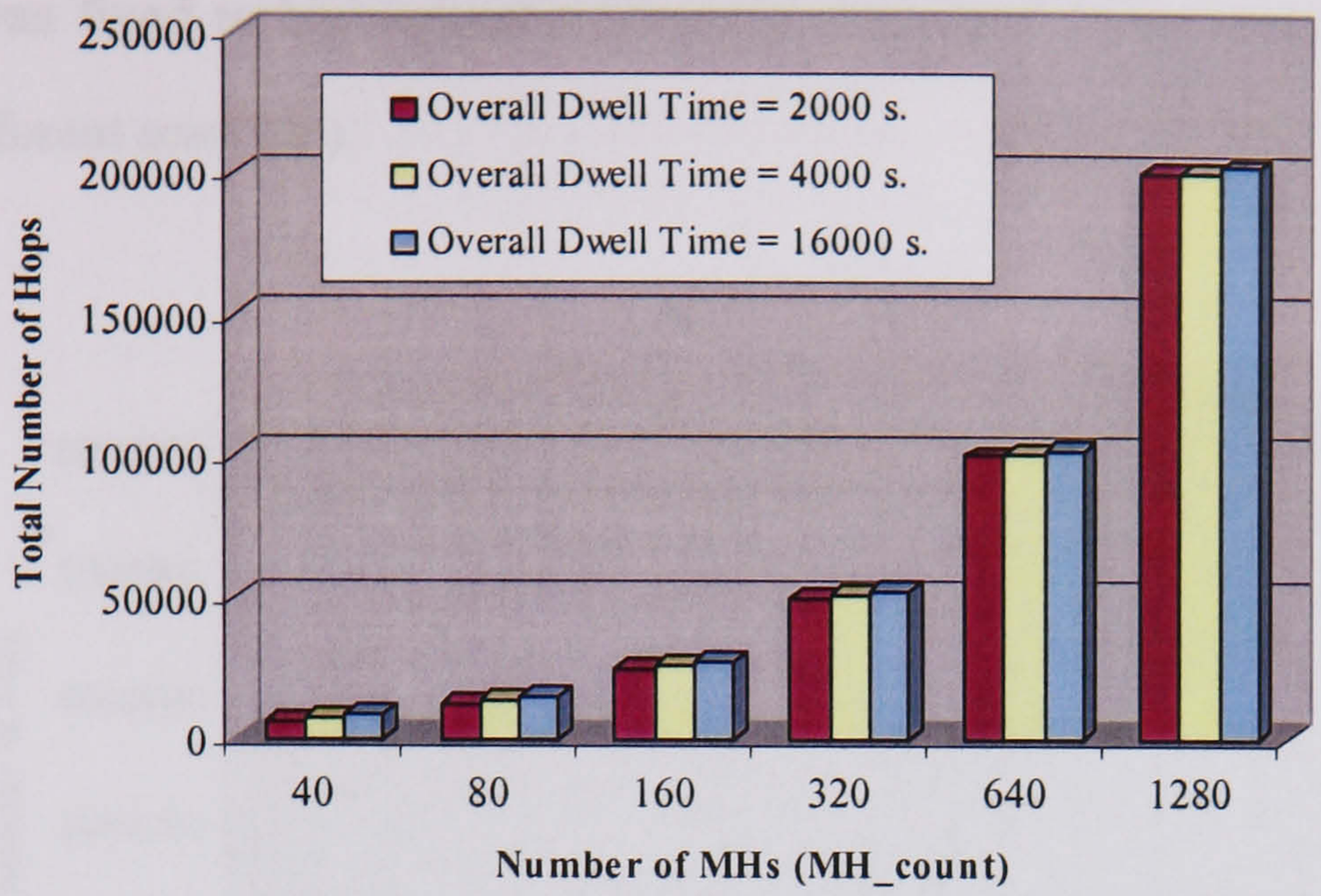
The first observation is related to the increased number of MHs in the network and differences in number of handovers performed in the network. This is shown in Graph 4.24 where $T_{overall}$ (overall dwell time) is fixed to 2000 seconds as applied in the simulations in section 4.3. MMP was analysed for the case where the constant MAX_RTX =1 and ECHO_INTEVAL =120 seconds following the case used in the simulations. There are 6 examples of different average dwell times ($T_{averege\_dwell}$) for MHs in the network taken from the simulation cases resulting in different numbers of handovers ($n_{handovers}$) performed during the $T_{overall}$: for $T_{averege\_dwell}$ =30 s. $n_{handovers}$=66, for $T_{averege\_dwell}$ =50 s. $n_{handovers}$=40, for $T_{averege\_dwell}$ =70 s. $n_{handovers}$=28, for $T_{averege\_dwell}$

=90 s. $n_{handovers}$=22, for $T_{averege\_dwell}$ =110 s. $n_{handovers}$=18 and for $T_{averege\_dwell}$ =130 s. $n_{handovers}$=15.



Graph 4.24. MMP: effects of increased population of MHs and different number of handovers performed

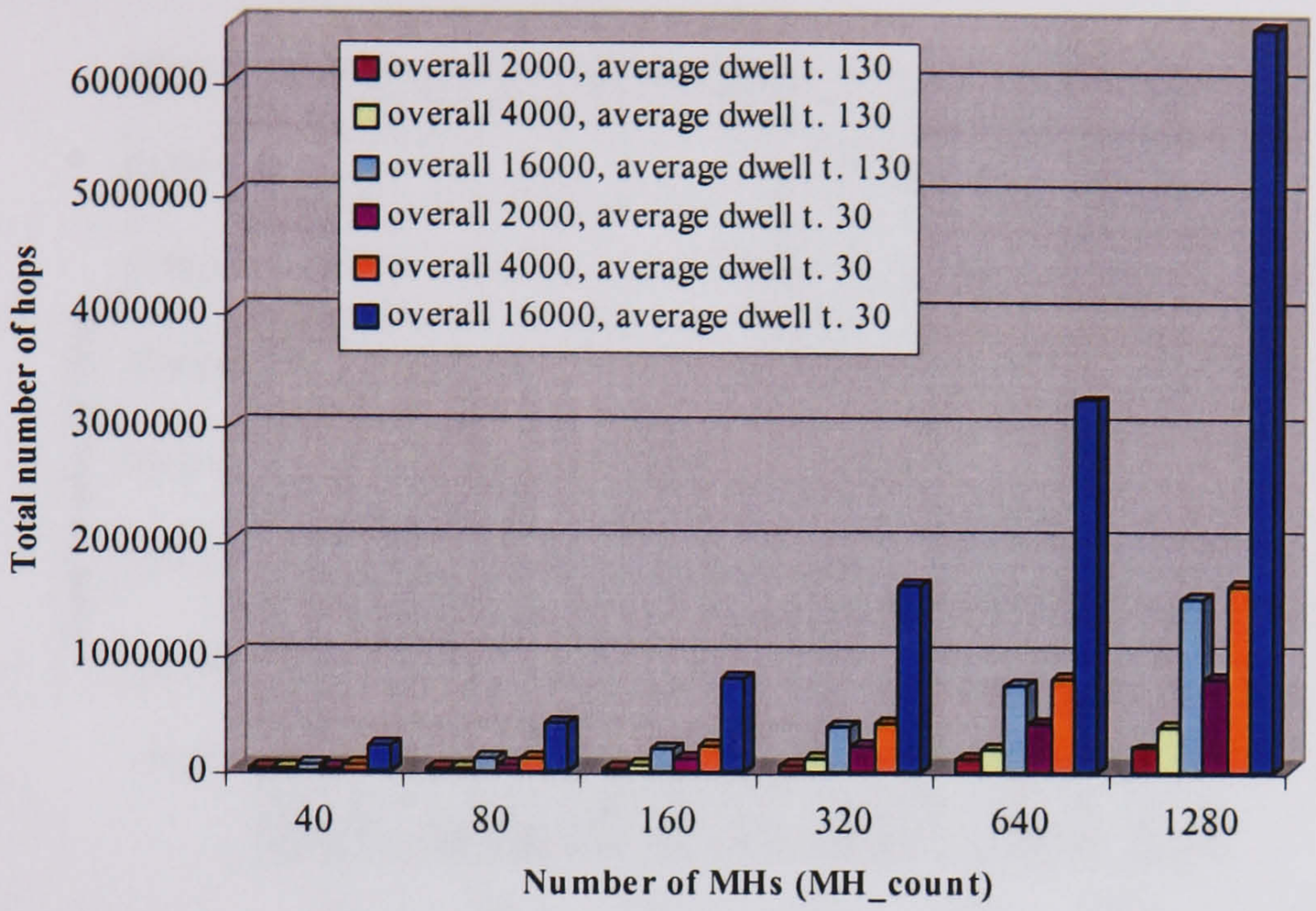
The scenario shown in Graph 4.24 shows that handovers, that is, movement-triggered messages account for most of the protocol overhead for larger populations cases. This was expected since the soft-state part (equation 4.26) of the equation 4.27 was fixed since it is independent on the number of MHs in the network and $T_{overall}$ and $T_{refresh\_interval}$ were kept fixed (for this case the soft state part of the total hops count is 333.3 hops). Example values for the most number of handover performed when $n_{handovers}$=66 can be taken for *MH_count* =640 and 1280. These are 408653.2 and 816973.3 respectively where the former is almost exactly the double of the latter value indicating that the movement-triggered messages account for most of the overhead in these cases. This is slightly less apparent for *MH_count* = 40 and 80 for $n_{handovers}$=15 with values of 6566.06 and 12798.79 respectively.



Graph 4.25. MMP: effect of increasing Overall Dwell Time for determining the impact of soft-state messages ($n_{handovers}$=15, fixed)

Based on the previous conclusion another case is examined where the focus was on the effect of soft-state messages in the overall hop count for MMP. This is intended to show the properties of multicast routing in MMP and aggregation of soft-state mechanisms. In the scenario shown in Graph 4.25 the same MMP setup was used as for Graph 4.24 where the overall dwell time $T_{overall}$ was varied taking values of 2000, 4000 and 16000 seconds. Movement of MHs was adjusted so MHs always perform 15 handovers (i.e. $n_{handovers}$=15, fixed) for every case shown in Graph 4.25: for $T_{overall}$ = 2000 s. $T_{averege\_dwell}$ = 130 s., $T_{overall}$ = 4000 s. $T_{averege\_dwell}$ = 260 s. and $T_{overall}$ = 16000 s. $T_{averege\_dwell}$ = 1040 s. Compared to the case when overall dwell time is 2000 seconds the cases of 4000 and 16000 seconds (increase of the overall dwell time by 2000 and 14000 seconds respectively) introduces 333.33 and 2333.33 more hops traversed by the soft-state message respectively as given by equation 4.26. The effects of using multicast CBT soft-state mechanism for MMP can be evident from the Graph 4.25 since the differences are the same for all populations of MHs and hence constitute a small proportion of the overall overhead since in the case of overall dwell time of 16000 seconds the quadruple increase in the overall dwell time brings the hop count increase by the factor of 1.012 for the case of 1280 MHs (Note: number of

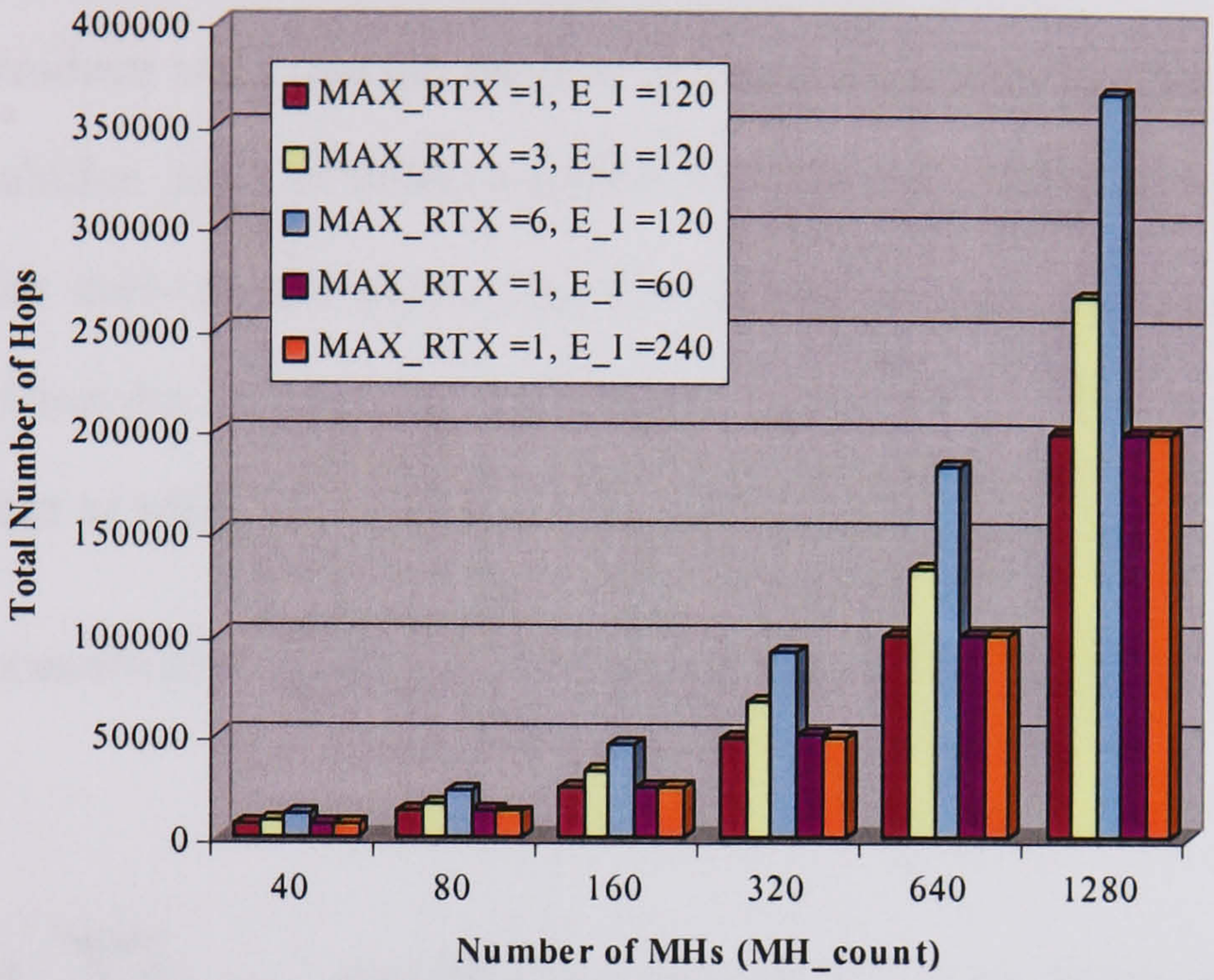
handovers was fixed to highlight this property associated with multicast, next graph applies a different scenario).



Graph 4.26. MMP: effects of varying overall dwell time while keeping the average dwell time in the cell fixed

In order to collectively analyse the properties associated with movement-triggered and soft-state messages shown in the previous two graphs, Graph 4.26 shows the effects of increasing the overall dwell time ($T_{overall}$) while keeping the average dwell time in the cell ($T_{averege\_dwell}$) fixed for the same scenario of MMP used in the previous two graphs. Therefore in this case, increase in the overall dwell time results in larger number of handovers performed (unlike Graph 4.25 where this was reversed since $n_{handovers}$ was fixed and $T_{averege\_dwell}$ was accordingly changing). The results of Graph 4.26 can be related to the properties observed in Graph 4.24 where the increase in number of handovers accounts for the large increase in hop count. The same situation is occurring in the case of Graph 4.26 since increase in overall dwell time is inducing more handovers as the average dwell time was fixed. One example is values when $T_{overall} = 2000$ s. and $T_{averege\_dwell} = 30$ s. and when $T_{overall} = 4000$ s. $T_{averege\_dwell} = 30$ s.

where the latter is almost the double of the former value (816973.3 and 1628128 hops respectively).



Graph 4.27: MMP: effect of varying CBT multicast constants MAX_RTX and ECHO_INTERVAL (E_I) in MMP

Since the protocol overhead of MMP is dependent on the values of CBT constant MAX_RTX and ECHO_INTERVAL ($T_{refresh\_interval}$), which are present in equation 4.27, Graph 4.27 shows the effects of varying both constants in MMP (note: ECHO_INTERVAL is abbreviated with E_I) for the case of $T_{overall}$ = 2000 seconds. and $T_{averege\_dwell}$ = 130 seconds considered in previous graphs. The first observation that can be deduced from the graphs and it is also evident from equation 4.27 is that the value of MAX_RTX influences the movement-triggered messages while the value of ECHO_INTERVAL ($T_{refresh\_interval}$) influences the soft-state part of the MMP hop count. Hence the same property can be observed as pointed out in the discussion on the previous graphs: movement-triggered messages account for the largest part of the protocol overhead thus liner increase of the values of MAX_RTX induces an almost exact proportional increase in the hops traversed. At the same time reduction or

increase of the value of ECHO_INTERVAL produces minor differences in the overall hop count (for ECHO_INTERVAL = 60 seconds 333.33 more soft-state related hops and for ECHO_INTERVAL = 240 seconds 166.66 less soft-state related hops compared to the case when ECHO_INTERVAL = 120 seconds).

In order to introduce and compare Mobile IP and Hierarchical Mobile IP with MMP for larger population cases of MHs, equation 4.6 used for validating Mobile IP results can be used for deriving the expression for Mobile IP and Hierarchical Mobile IP. Thus, the total number of hops for Mobile IP is given by the following equation for arbitrary number of MHs:

$$H_{(M.IP)} = MH\_count \times [2 \times \sum_{m=1}^{n_{handovers}} d_{(M.IP),m} + 2 \times (d_{max\_foreign} + d_{internet})] \quad (4.28)$$

where the part $\sum_{m=1}^{n_{handovers}} d_{(M.IP)m}$ signifies hop distance for every handover in Mobile IP and is fixed to the value of

$$n_{handovers} \times (d_{max\_foreign} + d_{internet})$$

since *Registration Requests* are sent to the HA. The part $2 \times (d_{max\_foreign} + d_{internet})$ is representing the login exchange of *Registration Request/Reply*. The new expression becomes

$$\begin{aligned} H_{(M.IP)} &= MH\_count \times [2 \times n_{handovers} \times (d_{max\_foreign} + d_{internet}) + 2 \times (d_{max\_foreign} + d_{internet}) \\ &= MH\_count \times 2 \times (d_{max\_foreign} + d_{internet}) \times (n_{handovers} + 1) \end{aligned} \quad (4.29)$$

Regarding Hierarchical Mobile IP the same basic approach can be applied as for Mobile IP shown in equation 4.28 where the expression for calculating hops traversed in Hierarchical Mobile IP is given by:

$$H_{(H.MIP)} = MH\_count \times [2 \times \sum_{m=1}^{n_{handovers}} d_{(H.MIP),m} + 2 \times (d_{max\_foreign} + d_{internet})] \quad (4.30)$$

Where the same assumption used in Mobile IP for the factor $\sum_{m=1}^{n_{handovers}} d_{(H.MIP),m}$ cannot be applied since the simulated setup of Hierarchical Mobile IP does not perform handovers with identical hops counts (see sections 4.2 and 4.5.1). The same was observed for MMP analysis for *Joint Requests* and *Join Replies* where the average hops count was introduced for calculating the hops count for any number of handovers. Hence for Hierarchical Mobile IP the new expression becomes

$$d_{average(H.MIP)} \times n_{handovers}$$

where $d_{average(H.MIP)}$ is the average handover distance for Hierarchical Mobile IP. The new final expression for Hierarchical Mobile IP can be represented as

$$\begin{aligned} H_{(H.MIP)} &= MH\_count\ [2 \times n_{handovers} \times d_a + 2 \times (d_{max\_foreign} + d_{internet})] \\ &= 2 \times MH\_count \times [n_{handovers} \times d_a + (d_{max\_foreign} + d_{internet})] \end{aligned} \quad (4.31)$$

As described in sections 4.2 and 4.5.1 for the simulated topology there are 8 possible handovers with *handover distance* of 1, and 3 handovers with *handover distance* of 3. Thus, as applied for MMP and for the simulated topology

$d_{average(H.MIP)} = 8/11 \times 1 + 3/11 \times 3 = 1.54545$

Performances of all three protocols are shown in Graph 4.28 where $T_{overall} = 2000$ and two cases of $T_{averege\_dwell} = 130$ and $T_{averege\_dwell} = 30$ resulting in $n_{handovers} = 15$ and $n_{handovers} = 66$ for all three mobility protocols (MMP: MAX_RTX=1, ECHO_INTERVAL =120.0 seconds). Internet hops are set to 1 (i.e. $d_{internet} = 1$). From Graph 4.28 it is evident that Hierarchical Mobile IP achieves the best performance as already noted in section 4.3. Since $d_{internet}=1$ Mobile IP performs better than MMP, although based on the conclusion from section 4.3, increase in the number of Internet

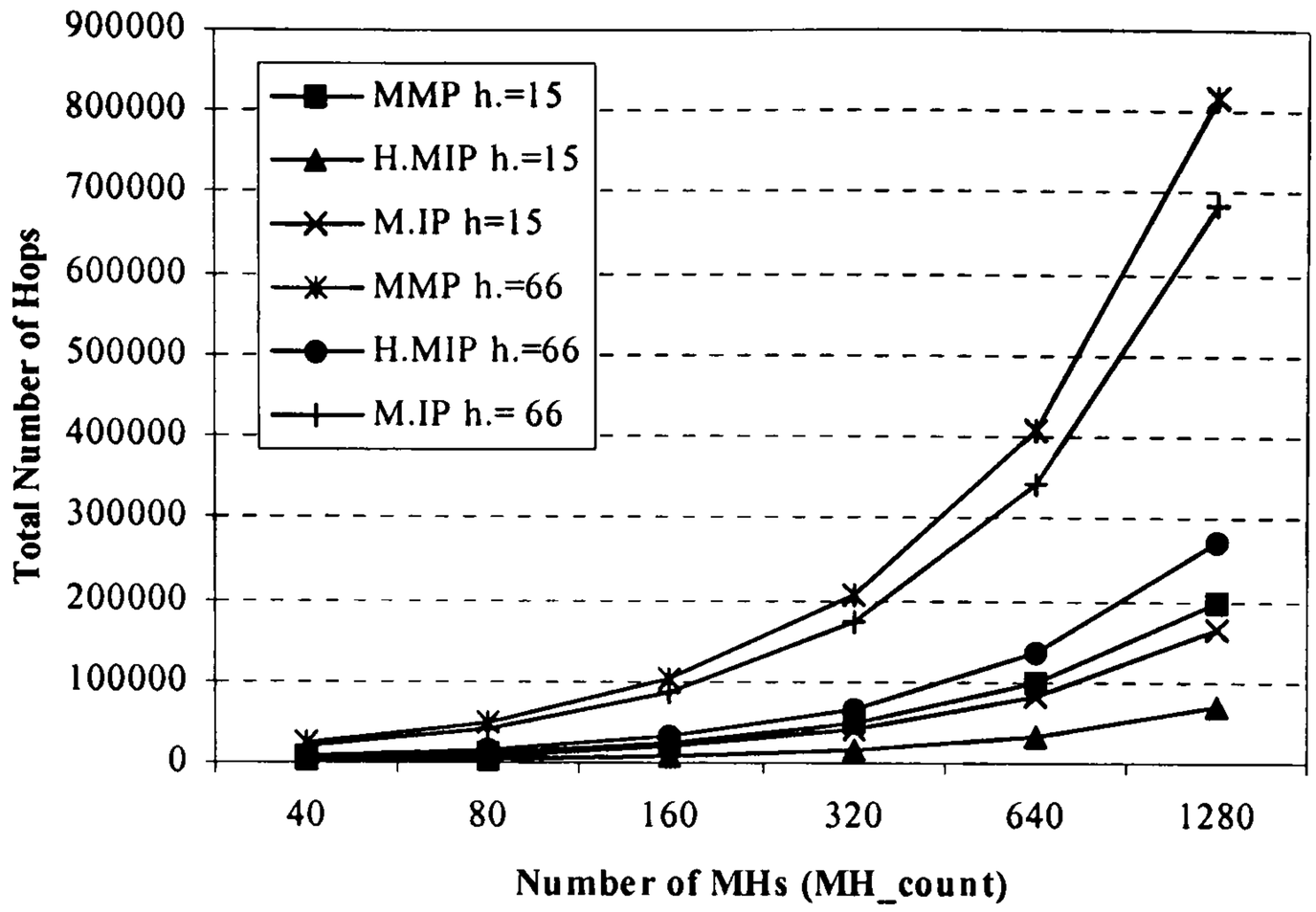
hops (i.e. $d_{internet}$) would results in Mobile IP performing worse than MMP (the effects of this are shown in new Graphs 4.16 and 4.17 and are not additionally considered).

Some of the conclusions from the MMP performance considering the previous graphs can be exemplified if two sets of values are taken for MMP and Hierarchical Mobile IP for small population cases of relatively slow moving MHs when $n_{handovers} = 15$ and *MH_count*= 40 (6566.06 for MMP and 2174.54 hops for Hierarchical Mobile IP) and for large population case of relatively fast moving MHs when $n_{handovers} = 66$ and *MH_count*= 1280 (816973.3 for MMP and 271360 hops for Hierarchical Mobile IP). Hence, the respective ratios between MMP and Hierarchical Mobile IP are 3.01952 and 3.01066 indicating that for larger population cases with more handovers performed the effect of soft-state messages are less apparent (although the differences are still significant and the improvements are minor).

In the following, the focus is on performing further comparisons between MMP and Hierarchical Mobile IP since Mobile IP remains the same for the simulated topology and the effects of increasing $d_{internet}$ (i.e. Internet hops) are already shown in section 4.3 as noted above.

As done for the analysis of handover performance in the previous section the model for calculating the hops traversed can be used for extending the analysis of Hierarchical Mobile IP for the case considered in the previous section where there is only two levels of FAs in the network this being the BSs and the Gateway: Modified Hierarchical Mobile IP as called in the previous section. Equation 4.31 can again be applied for determining the value of $d_{average(H.MIP\_modified)}$ for the simulated topology. In this case this calculation is simplified since modified Hierarchical Mobile IP performs identical handovers with *handover distance* of 3 thus giving

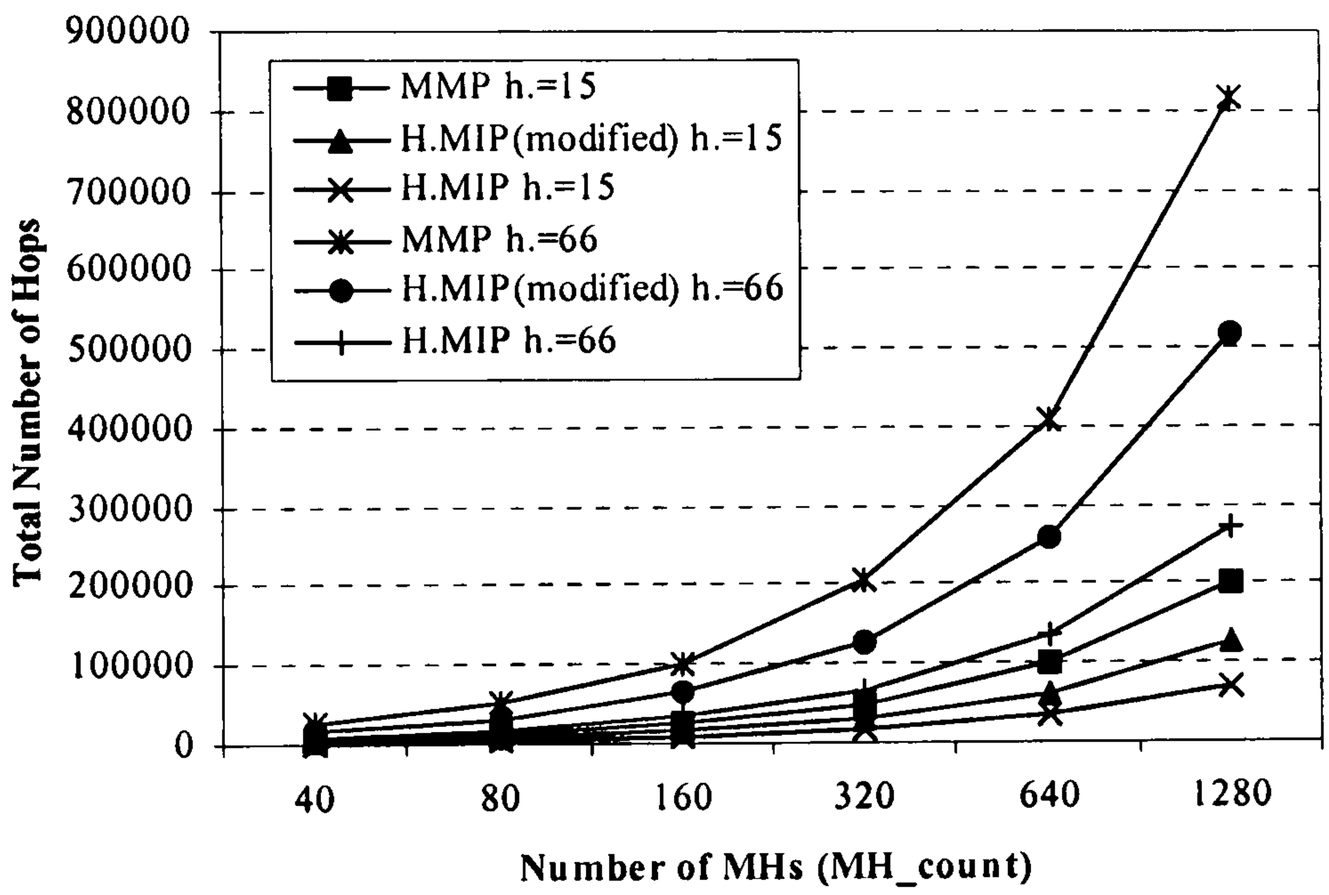
$$d_{average(H.MIP\_modified)} = 3$$



Graph 4.28: Performance of all three protocols for two cases of average dwell time

Equation 4.31 is applied for modified Hierarchical Mobile IP and compared to the results for MMP and Hierarchical Mobile IP shown in Graph 4.28. This is shown in Graph 4.29. The performance of modified Hierarchical Mobile IP is worse than the simulated scenario of Hierarchical Mobile IP. For the case examined in the previous graph when $n_{handovers}$ = 15 and *MH_count*= 40 (3920 hops for modified Hierarchical Mobile IP see above for the two other values for MMP and simulated-setup of Hierarchical Mobile IP) MMP induces 2646.06 more hops than modified Hierarchical Mobile IP and 4391.52 more hops than simulated-setup for Hierarchical Mobile IP (modified Hierarchical Mobile IP induces 1745.46 more hops that the simulated-setup of Hierarchical Mobile IP). For large population case of relatively fast moving MHs when $n_{handovers}$ =66 and *MH_count*= 1280 (517120 hops for modified Hierarchical Mobile IP see above for the two other values for MMP and simulated setup of Hierarchical Mobile IP) MMP induces 299853.3 more hops than modified Hierarchical Mobile IP and 545613.3 more hops than simulated-setup for Hierarchical Mobile IP (modified Hierarchical Mobile IP induces 245760 more hops that the simulated-setup of Hierarchical Mobile IP). The respective ratios between MMP and modified Hierarchical Mobile IP are 1.67502 for the case where $n_{handovers}$ = 15 and

MH_count= 40 and for the case where $n_{handovers}$ = 66 and *MH_count*= 1280 the ratio is 1.57985 (ratios between modified Hierarchical Mobile IP and simulated-setup of Hierarchical Mobile IP are 1.80268 and 1.90566 respectively where the ratios are the same for all values for the same number of MHs and same number of handovers). The differences in the ratios are due to the same properties observed in the previous graphs (although more apparent due to the higher values for modified Hierarchical Mobile IP), however, performance of modified Hierarchical Mobile IP is significantly worse that the simulated setup of Hierarchical Mobile IP.



Graph 4.29. Comparison of MMP with two cases of Hierarchical Mobile IP

The same strategy applied in the performance evaluation of handovers shown in the previous section can be demonstrated for the protocol overhead analysis where the models developed in this section can be used for performance analysis in the New Topology already applied in the previous section and shown in Figure 4.7. The new parameter for the New Topology is $d_{max\_foreign}$ = 4 for all three protocols ($d_{internet}$ = 1 is kept the same assuming one Internet hop between the foreign network and the MH's HA). Regarding MMP and the soft-state part of the equation 4.27 for calculating the

effective number of hops for MMP "keepalive" messages this is calculated as follows for the New Topology:

$$[\sum_{level=2}^{n} Node_{level}] - [Node_{(n-1)} \times (\frac{Node_n}{Node_{(n-1)}} - 1)] = 38 - 8 \times (3-1) = 22$$

where $n = 5$ and $Node_1 = 1$, $Node_2 = 2$, $Node_3 = 4$, $Node_4 = 8$ and $Node_5 = 24$.

Since there is 1 possible handover with *handover distance* of 4, 2 handovers with *handover distance* of 3, 4 handovers with *handover distance* of 2 and 16 handovers with *handover distance* of 1 the average *handover distance* for MMP in the new topology becomes:

$$d_{average(MMP\_new\_topology)} = (16/23 \times 1) + (4/23 \times 2) + (2/23 \times 3) + (1/23 \times 4) = 1.47826$$

The same can be calculated for the simulated-setup of Hierarchical Mobile IP with three levels of FAs as described in 4.1.2.1 and applied in previous section (also shown in Figure 4.7). Hence in the new topology the simulated setup of Hierarchical Mobile IP has 3 possible handovers with *handover distance* of 4 and 20 handovers with *handover distance* of 2. In order to apply this to equation 4.31, average *handover distance* is given by

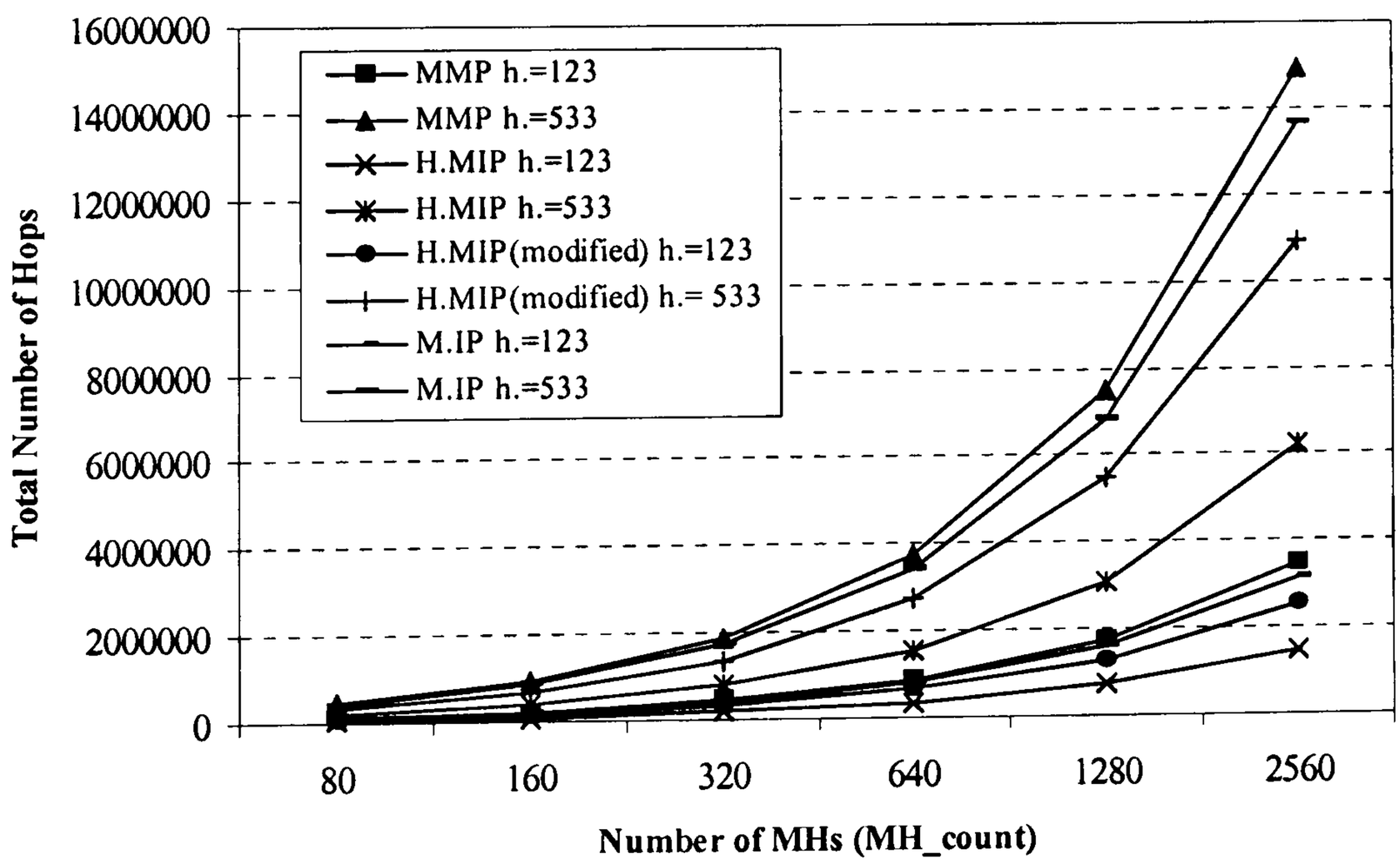
$$d_{average(H.MIP\_new\_topology)} = (20/23 \times 2) + (3/23 \times 4) = 2.26087$$

As done for the above analysis for the simulated topology shown in Graph 4.29 another scenario for Hierarchical Mobile IP can be introduced with only two levels of FAs this being the BSs and the Gateway (the same type of modified Hierarchical Mobile IP is applied for the handover performance in the new topology and in the previous graph for the simulated topology). Since the modified Hierarchical Mobile IP has identical handovers in the network, average *handover distance* is

$$d_{average(H.MIP\_modified\_new\_topology)} = 4$$

Performances of MMP (MAX_RTX =1, ECHO_INTERVAL =120 seconds), Hierarchical Mobile IP, modified Hierarchical Mobile IP and Mobile IP are shown in Graph 4.30 for New Topology where $T_{overall}$ = 16000 seconds and doubled population cases of MHs (i.e. *MH_count*) compared to the above analysis for the simulated

topology, for two previously applied cases of largest and smallest considered average dwell times: $T_{averege\_dwell}$ = 130 and $T_{averege\_dwell}$ = 30 resulting in $n_{handovers}$ = 123 and $n_{handovers}$ = 533 for all protocols. Similar behaviours can be observed as for the simulated topology case. For the case of relatively slow MHs with minimum population where $n_{handovers}$ = 123 and *MH_count* = 80 MMP has 114690.9, simulated-setup of Hierarchical Mobile IP has 45293.9 and modified Hierarchical Mobile IP 79520 hops traversed. When the result for MMP is divided with the result for simulated-setup of Hierarchical Mobile IP the ratio is 2.53215 and for when the same is done for MMP and modified Hierarchical Mobile IP the ratio is 1.44229. For the case of relatively fast moving MHs $n_{handovers}$ = 533 and *MH_count* = 2560 MMP has 14942571, simulated-setup of Hierarchical Mobile IP has 6195421 and modified Hierarchical Mobile IP has 10941440 hops traversed. When the result for MMP is divided with the result for simulated-setup of Hierarchical Mobile IP the ratio is 2.41187 and for when the same is done for MMP and modified Hierarchical Mobile IP the ratio is 1.36569 (note: the same properties of effect of soft-state message can be observed as for Graph 4.28 and Graph 4.29).



Graph 4.30: New Topology hop count for MMP and two cases of H.MIP

When these values are compared to the ratios for the simulated topology it can be concluded that due to the increase in network hops MMP manages to improve its performance relative to the other protocol because of the minimisation of *handover distances* and associated messaging. However, protocol overhead of MMP is still larger that the two cases of Hierarchical Mobile IP.

4.6 MMP Design Conclusions

4.6.1 Analysis of the Testing Strategy

One of the key objectives of the MMP design presented in this thesis is to show that the combination of Mobile IP and IP multicast can provide mobility support in the Internet and at the same time achieve comparable performances to other mobility solutions. MMP follows the logic applied in most of the recent mobility protocols proposed in the Internet research community (see classification, section 2.3.3) by focusing the operation of the protocol to scoped IP networks. In the case of MMP, this network was scoped by the Gateway (i.e. Core) and was referred to as the foreign network domain (or micro mobility domain). Testing of MMP was conducted through a similar simulation setup used for evaluating other IP mobility protocols as explained in 4.1.2 where the parameters used in the simulations are compared to other similar attempts in the outside work. The two different instances of network parameters, which define the *high-bandwidth* and *low-bandwidth* network cases provide indications of relative performances of the mobility protocols and comparative dependency on the network parameters used.

Feasibility of MMP's potential deployment is largely dependant on the complexity of the network topology where MMP may be implemented as a mobility solution. Assuming that MMP generally performs in a satisfactory way (the actual performance of MMP is further analysed in the remainder of this chapter but the previous sections indicate efficient performances concerning the handovers) in small topologies where

the scoped network consists of a small number of routers, the benefits of MMP could be suppressed because of the potential complexity of the protocol relative to the small deployment scenario. The complexity is assumed to come from the fact that all routers in the network are required to deploy CBT multicast protocol adapted for mobility in MMP. In these situations a subjective decision from the network operator may determine whether the deployment of MMP is feasible.

In order to understand this situation related to the deployment of MMP in relatively small networks, a comparison with *Proxy Agent Architectures* might be self-explanatory. A particular example can include a foreign network domain consisting of a single ingress/egress router (functioning as a Gateway for the network), which connects three "IP-routing-capable" BSs. By placing the top-level Proxy/Mobility Agent at the Gateway and bottom-level Proxy Agents at the BSs, mobility would be solved in a satisfactory way since it would be localised to the particular domain where minimum possible handover distances would occur for every possible intra-domain handover (intra-domain refers to all handovers between BSs belonging to the same network domain, see the next section for more details). At the same time, in cases of larger topologies MMP benefits are increased as shown in section 4.5.1.

The deployment feasibility of MMP is a general concern for all *Localised Enhanced-Routing Schemes*, because of the inevitable complexity induced by the requirement for improved efficiency of the protocols, which then results in more complicated protocol mechanisms. However, regardless of this trade-off (which may eventually be left to the preference of network operators), MMP would still perform efficiently even in small topologies as far as handover efficiency is concerned and would be easy to install because of the adaptable, dynamic and self-organising property of MMP's protocol mechanisms as they greatly rely on the available multicast routing protocol CBT which may already be present in the routing setup of the network. Finally, the current Internet research [47][48] and some of the outside work mentioned in 4.1.2 indicate that future IP-based networks will have topologies sufficiently larger than the

above considered small networks to fully justify the trade-off between the relative complexity of mobility protocols in small networks and potential performance benefits.

The models applied in section 4.4 for validating the simulation results and the additional handover performance analysis conducted in section 4.5.1 can be further extended for analysis of handover performances of the mobility protocols in networks with different transmission rates and link delays. In addition to different values for transmission rates and link delays, analysis in section 4.5.1 demonstrates how models could be adopted for a different network topology. This can be used to assist the analysis and provide some estimations of protocol performances in network scenarios not considered in this chapter based on the considered network parameters. Regarding the equation 4.5 used for calculating packet losses in section 4.4, this can provide a start for introduction of new network transmission rates and link delays. In the equation 4.5, transmission rates (r_n) and link delays (t_n) are expressed in $2\times\sum_{n=1}^{x}(\frac{S}{r_n}+t_n)$ giving the overall loop time for the particular *handover distance* x, which renders the packets loss when multiplied with the transmission rate of the traffic. The linear nature of equation 4.5 indicates that packet losses are expected to increase with the increase in link delays for every hop of the *handover distance* and that the increase in transmission rates of a link in the network, reduces the handover latency (i.e. handover loop time).

Simulations results support the properties of the theoretical model demonstrated in the performances of protocols for *high-bandwidth* and *low-bandwidth* network scenarios considered in the simulations. The *low-bandwidth* network has lower transmission rates and higher link delays. The consequence is that packet losses for the same traffic cases and *handover distances* are higher in the case of the *low-bandwidth* network as shown in section 4.2. This can be related to the theoretical model for calculating

packet losses given by equation 4.5 as transmission rates are reduced and link delays are increased for the *low-bandwidth* network.

Some more properties of packet losses are exemplified in Table 4.2 where *handover distance* of 3 (named *Reference Case* in the Table 4.4) is taken as an example situation and compared to different cases (*Case A* to *F*) with proportionally modified transmission rates and link delays. *Reference Case* is taken from section 4.4 where *handover distance* of 3 is validated against the simulation results for traffic rate of ω = 1.024 Mbits/s (for MMP and HMIP).

Case	Network Parameters	Description	Packet Loss
Reference Case	$_{l1}$ = 0.5 ms, t_2, t_3 =1.5 ms, t_4 =40 ms, r_1,r_2,r_3 =10 Mbits/s, r_4 =30 Mbits/s	*High-bandwidth* network parameters	14.614
Case A	$_{l1}$ =5 ms, t_2, t_3 = 15 ms, t_4 = 400 ms, r_1,r_2,r_3 = 10 Mbits/s, r_4 = 30 Mbits/s	Only link delays multiplied by 10	140.614
Case B	$_{l1}$ = 0.05 ms, t_2, t_3 =0.15 ms, t_4 =4 ms, r_1,r_2,r_3 =10 Mbits/s, r_4 =30 Mbits/s	Only link delays divided by 10	2.0144
Case C	$_{l1}$ = 0.5 ms, t_2, t_3 = 1.5 ms, t_4 = 40 ms, r_1,r_2,r_3 =100 Mbits/s, r_4 =300 Mbits/s	Only transmission speeds multiplied by 10	14.0614
Case D	$_{l1}$ = 0.5 ms, t_2, t_3 =1.5 ms, t_4 = 40 ms, r_1,r_2,r_3 = 1 Mbits/s, r_4 = 3 Mbits/s	Only transmission speeds divided by 10	20.144
Case E	$_{l1}$ = 5 ms, t_2, t_3 =15 ms, t_4 =400 ms, r_1,r_2,r_3 =100 Mbits/s, r_4 = 300 Mbits/s	Both delays & speeds multiplied by 10	140.245
Case F	$_{l1}$ = 0.05ms, t_2, t_3 = 0.15 ms, t_4 = 4 ms, r_1,r_2,r_3 =1 Mbits/s, r_4 =3 Mbits/s	Both delays & speeds divided by 10	7.544

Table 4.4. Example Cases of packet losses for different transmission rates and link delays

The first observation from Table 4.4 is that the applied values for link delays constitute the most important factor in determining the packets losses relative to the chosen *Reference Case*. This is shown when *Case A* and *Case D* are compared, where due to the changed parameters for the case of link delays (increased by 10) and transmission rates (reduced by 10) packet losses are expected to increase in both cases. However, increase of link delays produces larger increase in packet losses, as it constitutes the main factor in equation 4.5. The same property can be observed for *Case B* and *Case C* where packet losses are expected to drop in both cases compared to *Reference Case*. In *Case C* only minor reductions are occurring since the factor related to transmission rates is almost negligible compared to link delays. *Case E* and *Case F* further support properties of the previous cases where in *Case E* link delays

are an overwhelming factor for determining packets losses which are almost the same as in *Case A*. In *Case F* reduction in transmission speeds (beside the reduction in link delays) does not incur significant packet losses compared to the *Reference Case*.

A reflection can be made based on the simulation results and the theoretical models for calculating packet losses regarding the handover performance of the mobility protocols: the concept of *handover distance* and its additional application for calculating overall and average packet losses, indicates that the most successful protocol in terms of the handover performance is always the one that manages to minimise the *handover distance* for considered number of handovers. The extent and impact of this property based on the simulated scenarios is one of the topics of section 4.5.1 where MMP achieves the best performance. In scenarios with different network parameters, the extent of differences between the considered protocols would be related to the parameters for each network hop in the *handover distance* calculation.

4.6.2 Critical Analysis of the Simulation Results

The simulation results of this chapter are shown for the three protocols which belong to the three essential approaches in solving IP mobility: the basic Mobile IP and the one example from two other categories of Regional mobility protocols: MMP for *Localised Enhanced-Routing Schemes* and Hierarchical Mobile IP as the prime example of *Proxy Agents Architectures*. MMP simulation results and further performance analysis in section 4.5 confirm that MMP manages to significantly reduce *handoff latencies* (thus packet losses) mainly due to localisation of the updates of routing entries. In fact, MMP minimizes the *handover distance* to the smallest possible value for every handover case tested in a hierarchical topology, thus acting as the best possible solution as far as handover efficiency is concerned (this performance could also be achieved with some other mobility protocols specifically *Localised Enhanced-Routing Schemes*).

According to the simulation results for handover performance MMP is superior to Mobile IP and better than Hierarchical Mobile IP. As explained in section 4.2, the simulation setup of Hierarchical Mobile IP is such that packet losses for some handovers are identical to MMP for cases when the *handover distance* in MMP was one and three. For the MMP case when *handover distance* is two, Hierarchical Mobile IP performs a handover with *handover distance* of three due to the positioning of FAs/Proxy Agents in the network. Another scenario of Hierarchical Mobile IP is introduced in section 4.5 where there are only two levels of FAs in the network, in which case performance of MMP is further improved. Observing the performance of Hierarchical Mobile IP the handover performance could be improved if all routers in the network were configured as FAs/Proxy Agents resulting in optimal *handover distances* for every handover and causing Hierarchical Mobile IP to perform identically as MMP. Although this may be possible from the implementation perspective, it does not present a realistically comparative scenario since it would create an unrealistically complex structure of FAs. The potential benefits of using Hierarchical Mobile IP are very doubtful if every router in a network needs to function as a Proxy Agent of the mobility protocol as this would require instalment of needed features in every router in the network for any topology scenario (an the resulting requirements such as encapsulation/decapsulation of packets). Thus, it can be explicitly stated that MMP performs better than Hierarchical Mobile IP when considering only efficiency of handovers. In fact, based on the above discussion this could be extended to a general statement that *Localised Enhanced-Routing Schemes* are better candidates for achieving minimum *handover latencies* than *Proxy Agent Architectures*.

As indicated in the earlier parts of this document, *handover latency* (i.e. packet losses) is not the only factor in determining the suitability of a mobility protocol. Simulation was therefore used to obtain protocol overhead as another important efficiency parameter. Simulation results for the protocol overhead of MMP reveal a shortcoming

of MMP, also relatable to other *Localized-Enhanced Routing Schemes*: distributed and dynamic location management and fast route updating inevitably incurs an increase in generation and distribution of control messages, i.e. complexity induces protocol overhead.

It should be noted that much of the simulation results for MMP are affected by the complexity and protocol overhead-inducing procedures in bus links explained in section 4.3. While the introduction of bus links was believed to provide an insight into some particular situations of interest and presents an important testing scenario, their presence significantly increases the protocol overhead of MMP compared to other protocols as can be observed in the validation of protocol overhead in section 4.4 and further protocol overhead analysis of MMP in section 4.5.1 (more control message are used for handling bus link situations).

This point highlights one of the main aspects of performances and general effectiveness of the mobility protocols considered in the analysis: **the almost inevitable trade-off between the protocol's performance during handovers and the protocol overhead induced as a consequence**. This aspect is one of the foundations for the research presented in the following chapters.

Regarding the analysis of the protocol overhead in this chapter there are two aspects related to the increase of the protocol overhead in MMP and alike solutions:

- the actual number of control messages a mobility protocol needs to complete an operation and,
- the amount of resources these control messages actually consume, expressed in terms of the number of hops they traverse.

The traversed hops are used for further analysis of MMP and the other two tested protocols shown in section 4.5.2. They essentially replace a non-dimensional protocol cost analysis with the difference that they reflect an architectural (i.e. topology

related) metric useful for indicating the differences in protocol costs relative to the considered scenarios.

The analysis could be practically extended beyond the applied architectural indicators by using the calculations in section 4.5.2 for injecting messages sizes and observing the practical cost the control aspects for each protocol (this would have to take into account some specific implementation issues such as IP versions and protocol versions).

Another issue for consideration can be defined by the following question: is it more desirable to have a comparatively small number of control messages traversing a large number of routers where some of them are in the global Internet (Mobile IP case) or to have more messages traversing scoped environments such as the foreign network domain scoped by the Gateway (this is also a general *micro mobility* case). This question sometimes induces a subjective answer but, certainly, from a global point of view of general Internet development, the second option is more acceptable since the consumption of local network resources is up to the local network operator and can be controlled by the scope and type of mobility support for the visiting MHs.

Simulation results reveal that the MMP protocol overhead is essentially larger than the other two protocols tested. However, since most of the protocol mechanisms of MMP are local (in the scoped network), the resulting consumption of the overall network resources is reduced. It can be deduced that the average number of hops traversed by a single message is less than in Mobile IP, but still not as efficient as Hierarchical Mobile IP. Section 4.3 includes experiments with adjustments of protocol constants of MMP to decrease the density of control message generation. The simulation results obtained show a significant reduction in the overall protocol overhead (around 20000 fewer messages and hops for a case with the most dense population of users). Another factor in the evaluation of MMP is the distance between the Gateway and the HA in the *macro mobility* section of the protocol. If the number of hops separating the Gateway and the HA is increased it can lead to situations as

shown in Graph 4.16 where Mobile IP messages, although fewer in number, traverse more hops than the control messages of the modified MMP. The same principle would apply in favour of MMP if the hops in the foreign network were further reduced.

Hierarchical Mobile IP theoretically produces the smallest protocol overhead for all the different cases of network setups. This was generally expected, since the setup of Hierarchical Mobile IP generated the same number of message as Mobile IP and reduced their "journey" through the Internet due to the hierarchical structuring of FAs, which terminate the Registration Requests and acknowledge them with Registration Replies. The example additional scenario of Hierarchical Mobile IP, called modified Hierarchical Mobile IP, used in section 4.5.2 shows that the protocol overhead of MMP approaches the one of Hierarchical Mobile IP especially for larger topologies. As already pointed out, simulations of Hierarchical Mobile IP and Mobile IP do not include the effects of refreshments (these could be in the order of minutes), which are expected to be an implementation feature common to all protocols.

Although not shown in the simulations of Hierarchical Mobile IP, the presence of so many FAs (i.e. Mobility Agents) does itself present an overhead due to the extra features they require in order to operate and process the packets. The simulation results and further performance analysis fail to show this property. As a comparison with MMP, this property is a significant drawback of Hierarchical Mobile IP since MMP mostly uses default and adaptable routing methods (i.e. multicast) available in the Internet. Hierarchical Mobile IP is not adaptable to different network scenarios because the positioning and instalments of FAs is manual and has to be performed by network operators. This is especially important in a comparison with MMP, which is a dynamic protocol based on the multicast routing, and is expected to "fit into" any network environment and points at the non-quantitative parameters that could be considered when comparing the protocols.

Another point which needs to be considered when comparing Hierarchical Mobile IP and all *Localised Enhanced-Routing Schemes* (including MMP) that the current state

of research in the Internet has almost distinguished the two types of solutions where Hierarchical Mobile IP is being proposed as a supplement to Mobile IP and the research into micro-mobility solution (*Localised Enhanced Routing Schemes*) is ongoing as a separated issue[10]. This further enforces the validity of the MMP's design and introduces the issues discussed in the following as the main criteria for evaluation of MMP.

MMP is still a very promising solution for solving IP mobility despite its protocol overhead. Compared to Mobile IP, MMP is certainly more appropriate since the handover performance is significantly improved and the protocol overhead escalation, with some adjustment of the protocol constants, is acceptable (for some cases of locations of HAs relative to the foreign network, MMP is more efficient even in terms of the protocol overhead). However, the question of efficiency of Hierarchical Mobile IP is not so much a simulation topic but is the extent of its suitability as a ubiquitous mobility solution.

The study of protocol overhead in the conducted simulations is a starting point for consideration of scalability of MMP and other protocols. Number of MHs in the simulations was limited to twenty. While this population of MHs can be considered a realistic approximation for the tested types of networks (i.e. typical local area networks/sites) mobility protocols are also expected to be deployed in significantly larger networks with larger and denser population of users/MHs. The simulation results show behaviours of protocols with the gradual increase in the population of MHs (from one to twenty). This is assumed to give an indication of how population increase affects the protocol overhead, thus, indicating how the protocol scales with the increase in number of MHs. The main concern with all *Localised Enhanced-Routing Schemes*, including MMP, is the trade-off of the improved handover

[10] In fact at the time of the writing of this thesis, Hierarchical Mobile IP and Mobile IP are being engineered in the IETF Mobile IP Working Group, while the research into micro mobility protocols (i.e. *Localised Enhanced Routing Schemes*) has been "shifted" to Internet Research Task Force (Micro

performance at the expense of increased complexity and implications of this on the scalability of solutions. This was the primary reason for additional protocol overhead performance analysis of MMP in section 4.5.2 for significantly larger population cases of MHs.

Before the scalability of MMP is further analysed an overview of the general meaning of scalability can be useful for understanding the entire set of issues that play parts in the overall assessment. Thus, the following issues can be related to the scalability of a mobility protocol in addition to the performance metrics mentioned in section 4.5.2:

1. Generation (number of control messages generated) and distribution (hops traversed by the messages) relative to the increase in population of MHs in a network.
2. Generation (number of control messages generated) and distribution (hops traversed by the messages) relative to the increase in density of MHs in a network (e.g. number of MHs per cell).
3. Correlation of the above two for different network sizes and user behaviours (speeds, traffic patterns,…).
4. Impact of the population increase on the Internet connectivity. This point mostly related to address managements, i.e. how does the increase in population of MHs and allocation of addresses affect to affect the address managements in a network and outside.
5. Increase in other aspects of protocol overhead: processing, tunnelling overhead, requirements for special-purpose routers (Mobility Agent, FAs, Gateways…).

Regarding the performance of MMP the simulation results and the validations show that the protocol overhead does not linearly increase with the increase in population of MHs (as with other protocols) in the population cases considered in the simulations. This aspect can be mostly associated to the aggregated “keepalive” mechanisms in

---

Mobility Routing Areas Subgroup). Also topic of some of the parallel research activities, which influenced research shown in next chapters

MMP (*Echo Reply* and *Echo Request*). In the additional analysis in section 4.5.2, this property is explained by extracting the soft-state messages, which are aggregated for all MHs using a particular routing branch. However, another property was observed for large population cases and it is related to the effects of "keepalive" messages (i.e. soft-state messages). They account for a small portion of the protocol overhead where mobility related signalling becomes a dominant factor with almost linear effect on the protocol overhead increase.

The number of control messages generated in MMP is always expected to exceed the ones of Mobile IP and Hierarchical Mobile IP due to the more complex handover procedures. However, the simulation results also reveal that although larger in number, control messages in MMP can traverse less hops collectively that the ones in Mobile IP due to their localised distribution. This is an important scalability advantage of MMP and it is expected that this aspect would be further highlighted in situations with large number of MHs and large networks where the distance between HA and the MH's current network increases (i.e. increase in Internet hops).

Additionally the setup of MMP is such that the address management does not impose scalability risks since the addresses are managed locally and are transparent to outside of the foreign network as discussed in Chapter 3.

As already mentioned, the overhead of Hierarchical Mobile IP can also be seen in terms of the installed Mobility Agents and the associated processing (decapsulation/encapsulation, processing of messages…) which for large and dense population of users and large networks can be a significant factor as it would consume a considerable amount of network resources. While the five scalability points can easily be associated with any *Localised Enhanced-Routing*, MMP has a particularly beneficial property for large and more specifically dense population of users. The "soft state" mechanism of MMP presents a major advantage of MMP's mechanisms for maintenance of multicast routing entries due to the aggregation of "keepalive"

messages (*Echo Request* and *Echo Replies*) for every hop irrespective of the number of MHs using the particular hop (i.e. tree branch) from the Gateway to the serving BS. The remaining paragraphs deal with the problems of MMP's deployment and a comparison with some other mobility schemes.

Pre-MMP Multicast-based schemes concluded their proposals with common conclusion that the current Internet is not entirely adjusted to meet the requirements for full scale, efficient multicast for solving mobility (see section 3.1.1 and [38]). This includes lack of multicast capable routers and the necessary modification of the TCP/IP protocol suite. Although the setup of MMP overcomes some of the previous problems such as the TCP support, the problem the global availability of multicast capable routers (especially those running CBT) is relative to the ongoing global effort of Internet development and is expected to be resolved in the near future. As an example, the architecture/topology dependency is hardly an issue since the globally accepted concept of *micro-mobility* bases its design on architectures similar and not less complex than the one used as a test network in this document. Additionally, the ongoing processes of development of telecommunications systems rely on the fact that new communication devices, including IP routers, will be equipped with the ability to support all demanding technologies such as multicast. This further enforces the flexibility of MMP since, unlike some other mobility protocols, it bases most of its mechanisms on the existing Internet protocols, IP multicast and Mobile IP, which are highly likely to be features of all future IP networks.

Compared to the existing Regional mobility protocols, MMP stands as a very efficient protocol from several aspects. As mentioned for the case of Hierarchical Mobile IP, unlike the *Proxy-Agents Architectures*, MMP is not dependent on the pre-configuration of Mobility Agents. In order to make the setup of any *Proxy-Agents Architectures* as efficient as MMP during handoffs, every router in the network has to be configured as either a Proxy Agent or a FA. This would create expensive and highly inflexible network architectures and the whole concept of such a protocol will

not provide a generic solution for the emerging networks. Contrary to this, MMP adapts to any network topology without the risks of looping (basically the design of CBT is proven to avoid any routing loop due to the tree forming procedures [28]). Although tested in a typical hierarchical topology, there are no limitations to its successful performance in other network configurations including a mesh topology as the created multicast tree is expected to always form an effective hierarchical topology since the tree forming message are addressed to the Gateway/Core (assuming shortest path routing to the Gateway). Performance of MMP during handoffs can be assumed very similar to other examples of *Localised Enhanced-Routing Schemes* such as Cellular IP and HAWAII as they apply the concept of per-host based forwarding using a preinstalled routing tree as done in MMP using multicast (and the consequent updating of the "cross over" router resulting in minimised *handover distances*). One advantage of MMP compared to the two schemes is that MMP is based on the existing CBT protocol as the *enhanced-routing* scheme. The required modification of protocol stacks in the base station and the Gateway/Core and the addition of a new *MMP Instruct* message seem minor compared to the requirements of the other two schemes which require deployments of completely new routing methods throughout the network.

Finally, MMP is designed as an "overlay" to Mobile IP and achieves this completely since the MHs are using the Mobile IP mechanisms unchanged and do not require any modification. This makes MMP a simple and flexible alternative to current mobility protocols, which can be realised easily in the Internet because it relies on existing protocols and does not require complex adjustments of all network entities.

4.6.3 MMP Research Conclusions

The previous sections concluded on the performance and design aspects of MMP and its application for solving mobility of Internet hosts. One of the conclusions from the conducted research is that there is a trade-off between achieving improvement of

handover performance and complexity associated with it. This was exemplified in the quantitative comparison of MMP with the other two considered protocols: Mobile IP and Hierarchical Mobile IP where MMP achieves the best handover performance at the expense of larger protocol overhead for the considered scenarios in the simulations. Based on these quantitative parameters and the particular preferences that may be placed on one of them, it cannot be explicitly stated that MMP is the most efficient candidate for solving mobility of IP hosts.

In addition to the quantitative parameters, the previous sections indicate some architectural issues that may be considered as factors for assessing the suitability of MMP as well as other mobility protocols. Some examples of these "qualitative" properties of the protocol are: use of multicast instead of protocol-specific routing installations in the network and elimination of any MMP-specific requirements on MHs as they are assumed to execute Mobile IP where MMP-specific mechanisms are transparent (see section 3.4). The latter example is considered to offer deployment flexibility of MMP as it can always be realised as an option in the actual network ("on top" of Mobile IP). However, although considered as an important design decision, transparency of MMP-specific mechanisms to MHs, it actually induces more protocol overhead because of the features associated with this setup, i.e. *MMP Instruct* messages generated (see section 3.4.2, 4.2. and 4.5.2). In the alternative case where MHs are aware of the MMP-specific mechanisms multicast group management protocol (i.e IGMP) could be used for sending "leave" messages to the old BS to "cut off" the old routing branch and use of *MMP Instruct* message would be redundant.

The collection of the trade-offs based on different performances and qualitative architectural criteria and the associated variety of deployment and design-principle driven perspectives on suitability of each criteria, led to the research presented in the next chapter where alternative solutions are sought by considering various types of evaluation metrics.

CHAPTER FIVE

Generic Mobility Design Model

Chapter Overview

This chapter presents a novel method for evaluation and further development of IP mobility solutions using the experience from development of MMP and relevant mobility protocols. The chapter firstly proposes a method, called Evaluation Framework, for specific evaluation and analysis of IP mobility protocols. It then reflects the methodology and concepts of the Evaluation Framework (in particular Protocol Design Issues and their Solutions) by extending the analysis and comparison of MMP and some other mobility protocols. This is finalised in the concepts of generic mobility design model and impact of Protocol Design Issues. A particular application of the model is shown in the design processes applied in creation of mobility solutions in the BRAIN (and MIND) projects. This also includes analysis of BCMP, mobility protocol developed in the project, included for descriptive purposes due to its conformance to results of the model (BCMP is not author's contribution). Some of the conclusions of the presented model are further reflected against the default features of MMP.

5.1 Introduction

The conclusions on general efficiency of MMP, shown at the end of the previous chapter, significantly relate to two considerations. The first one is related to the quantitative performance of MMP and can be summarised in the performance trade off of increased protocol overhead for achieving more efficient handover performance. The second consideration relates to variety of qualitative considerations related to general features of mobility protocols for which evaluation processes are often descriptive and subject to deployment preferences (e.g. use of multicast routing for MMP). The example comparison of MMP, Hierarchical Mobile IP and Mobile IP shown in the previous chapter reveals the extent and types of comparative differences between the protocols. It can be generally stated that handover performance improvements, achieved by micro-mobility protocols, (in particular MMP and mobility protocols belonging to the *Localised Enhanced Routing Schemes* explained in chapter 2) induce various types and extent of protocol overhead. Reference to this statement is performance of Mobile IP and largely Hierarchical Mobile IP being an improvement of the basic features of Mobile IP (dependent on the placement and number of FAs in the network as discussed in the previous chapter). The experience from the design of MMP indicates that the protocol's overhead comes largely as a consequence of dynamic installations and adaptation of MHs' routing state in the network and the resulting maintenance of it. Chapter 2 contains an extensive classification of mobility protocols, which indicates a vast number of available proposals thus further complicating the decision about the efficiency of a particular solution. In addition, the Internet research community[1] is still in the process of discovering the most promising candidate for micro mobility, which suggests that no single solution has been recognised as the most successful.

The research presented in this chapter continues the work on mobility protocols by proposing a framework for comprehensive evaluation of their aspects and a model for construction of mobility protocols. The goal is to further expand on understanding of solving Internet mobility considering network environments applicable, and to consequently assist in selecting or developing suitable mobility protocols for specific deployment circumstances. The research is not intended to eliminate or negate the available experience in developing mobility solutions; rather, this is used as an input for the evaluation and development phases. This applies to MMP and its general properties noted in the previous chapter and general motivation for developing mobility protocols formulated in the initial design principles described in section 2.3.2. However, the following research offers models for constructive understanding and comparison of performances of mobility protocols including MMP. Specific features of mobility protocols (i.e. Protocol Design Issues) are promoted as modules that can be used for further development of mobility protocols in specific deployment scenarios and general understanding of applicability of their features.

The ultimate goal of the design model presented in this chapter is not a single and functionally ideal mobility solution, i.e. a single generic Mobility Management protocol. Rather, the objective is to devise a model for reaching a practical compromise between all proposed mobility protocols and for allowing further design attempts by presenting a platform for constructive input of design criteria and parameters, which can be subjective and deployment-specific.

The results of the research shown in this chapter are organised in the following way:

- A novel model for evaluation of mobility protocols is presented in section 5.2. The model is named Evaluation Framework consisting of Evaluation Criteria and Protocol Design Issues. In order to describe concepts of each Protocol Design Issue a description of their Solutions in some of the existing mobility protocols including

[1] At the time of writing of the thesis no micro-mobility protocol has been promoted as standard in the IETF. Recent efforts include creation of Micro-mobility research sub-group of IRTF/IETF for

MMP is given in section 5.2.2. This overview of available mobility protocols reflected against each Protocol Design Issue is also intended to provide a comparative analysis of the actual matching features of mobility protocols and the representing functionality of each Protocol Design Issue.

- Once each Protocol Design Issues is described the next step is extraction of all different types of Solutions for each Protocol Design Issue and indication of the general interdependencies between them. This is shown in section 5.3 constituting the main logic of the Generic Mobility Design Model by presenting the default interrelations between the Protocol Design Issues. Each Protocol Design Issue is described as a design driver, i.e. it is investigated as the dominant feature of mobility protocol design. Hence, the shown interdependencies can be used as a guideline for constructing the remaining parts of a mobility protocol.
- One application of the principles of the Generic Mobility Model is shown in section 5.4 extracted from development work conducted in the BRAIN project for creating mobility solutions. The section shows particular BRAIN design principles, resulting selection of BRAIN "primary" Protocol Design Issues constituting the main elements of the BRAIN mobility solution. One example mobility protocol is BCMP developed in the project shown as a mobility protocol conforming to the BRAIN mobility solution. As already noted, BCMP is not author's contribution but resulted from the work performed in the BRAIN projects (see section 1.6). The inclusion of BCMP helps to demonstrate a practical instance of the application of some of the design processes described in this chapter.

5.2 Evaluation Framework for IP Mobility Protocols

As already indicated, the first step that led to the design model for generic mobility solutions was the framework for productive evaluation of the mechanisms of IP mobility protocols. In order to analyse all mobility protocols and come to a conclusion

---

performing additional research on the topic.

on the applicability of their protocol mechanisms, a sophisticated method for evaluation of protocols could be applied [47][5][8]. This model for the evaluation of mobility protocols is named the Evaluation Framework and consists of two main elements:

1. **Protocol Design Issues (PDIs)**: These are the functional requirements for any IP-mobility protocol. Their characteristics can be deduced from the available mobility protocols. The feature of a mobility protocol, which solves a particular PDI, is called the PDI Solution. As the study detailed in the following text shows, PDIs and their Solutions have various levels of functional interdependences along the various functional axis of separation. The identified PDIs are: Packet Forwarding, Path Updates, Handover Management, Support for Idle MHs, Requirements for MHs, Requirements for Global Internet Interfaces, Address Management, Routing Topology and Security.
2. **Evaluation Criteria:** These are used to validate the effectiveness of a particular PDI Solution. Note: The evaluation criteria are a discrete set of requirements, derived from the initial design principles of mobility design outlined in section 2.3.2. While the initial design principles present an essential parameterisation of the default requirements for mobility support in IP networks, Evaluation Criteria are derived from the experience in analysing the actual features and performances of mobility protocols designed to tackle the initial requirements. Also, breaking down the Evaluation Criteria into distinct set of topics and their elements assists in fragmenting the problem into more obvious design challenges.

In other words, the basic idea is to firstly note what a protocol must be able to do, and then an assessment should be made on how well the protocol performs. The Evaluation Framework can benefit from the functional classification of mobility protocols presented in Chapter 2 to initially perform a top-level analysis of the protocol classes and to then to concentrate on the separate protocols from candidate classes. In fact, one of the first objectives of the Evaluation Framework, applied in the

BRAIN project, was to assist in narrowing down selection of suitable mobility protocols. Initially, intention was to start with the categories of mobility protocols and to use the evaluation criteria to make final judgement on suitability of each category (see section 2.3.3 for classification of mobility protocols) and to then repeat the same process for each protocol in the chosen category in order to select the most suitable protocol. Consensus was not reached on applicability of a specific existing mobility protocol (section 3.6.1 in [47] and [8]), hence the specific mobility results are proposed.

The Evaluation Criteria is presented first in the next section after which the PDIs are explained in more details and their Solutions showed by reflecting the corresponding features in some exemplar mobility protocols including MMP.

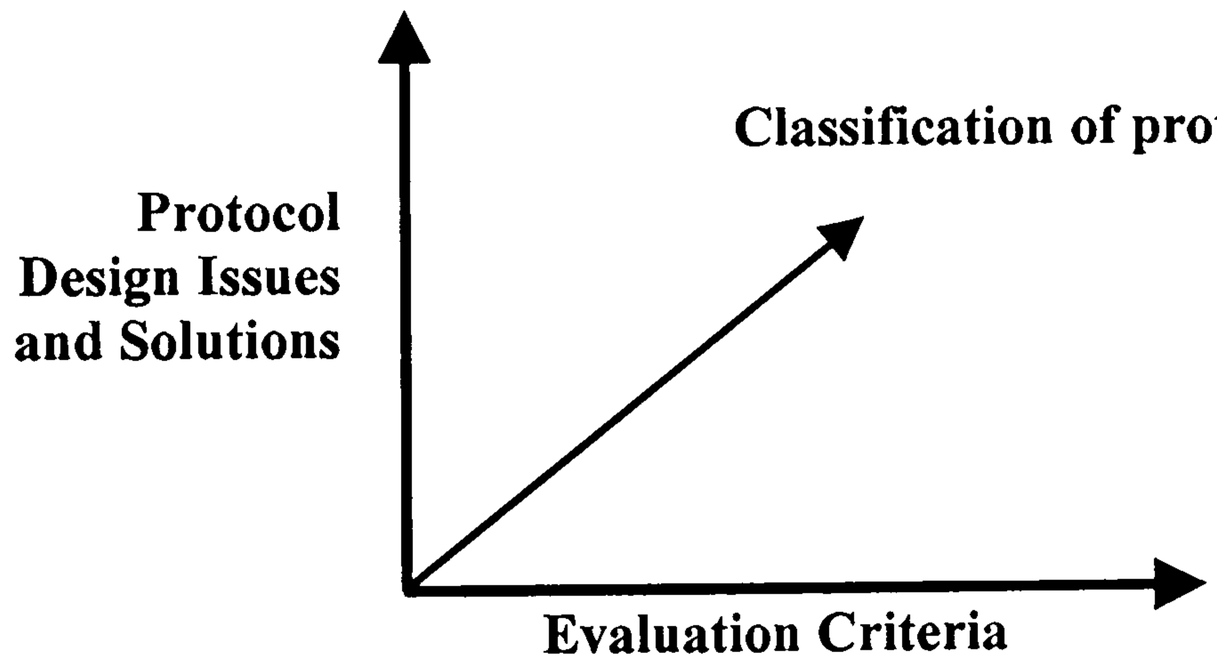


Figure 5.1. Evaluation Framework Concepts

5.2.1 Evaluation Criteria

The second part of the Evaluation Framework identifies the Evaluation Criteria. These can be grouped into 3 broad topics:

a) Efficiency

i. minimal end-to-end packet delays

ii. minimal handover latency i.e. no significant packet loss, reordering or duplication

iii. good throughput

iv. optimised routing (including the MH-to-MH case as shown in section 3.4.3.4)

v. small signalling load over wired and wireless links i.e. small protocol overhead

b) Scalability and robustness

i. support of a large number of MHs with different behaviours

ii. support of any population of routers in a domain

iii. support of a large amount of traffic per MH

iv. resistance to extreme cases such as link or router failures, i.e. no single points of failures, wireless links errors

v. resistance to routing loops

c) Applicability/Ease of deployment

i. simplicity

ii. compatibility with the standard Internet protocols

iii. ability to support Int-Serv/Diff-Serv QoS protocols [3]

iv. ability to support dumb MHs that are Mobile IP compliant

v. ability to adapt to changes in the network topology

vi. applicability of the same basic approach to both IPv4 and IPv6

The idea behind the identification of the above topics and their elements is that most of the more specific design and deployment parameters can be extracted from them (it should be noted that there is a possibility that few more elements may be added to the identified set as the research may evolve further). Then a particular sub-set of the Evaluation Criteria, or idealistically the entire set, can be used as the key design and evaluation metric.

5.2.2 Protocols Design Issues and some exemplar PDI Solutions

The following sections present an analysis of each Protocol Design Issue (see Figure 5.2) and some PDI Solutions, comparing four exemplar protocols and drawing out points of interest. The analysis is qualitative, thus little is said about the "efficiency" criteria, which is largely quantitative and which is exemplified in Chapter 4 in the study of handover latency, i.e. packet losses, and protocol overhead. The exemplar protocols are chosen from four different classes of mobility protocols identified in Chapter 2: one from *Proxy Agent Architectures* (Hierarchical Mobile IP[2]), and one from each of the sub-classes of *Localised Enhanced-Routing Schemes*: HAWAII from *Per-host Forwarding Schemes*, MMP from *Multicast-based Schemes* and MER-TORA from *MANET-based Schemes*. There is a commonality between the concepts and functions represented by the Protocol Design Issues and the abstract mobility model explained in section 2.3.2. While the latter presents the essential conceptual mechanisms required to support movements of IP hosts from a perspective of default Internet Protocol functionality, the former indicates a more detailed set of practical functionalities, which can be found in almost all mobility protocols. In other words, Protocol Design Issues and their solutions (PDI Solutions) are the actual realisation of the broad concepts laid out in the abstract mobility model. Additionally, PDIs contain some additional features, which have became evident during the course of IP mobility research and are not necessarily evident from the pure mobility perspective of the abstract mobility model.

The extent of the elaborated PDIs and their Solutions in four chosen protocols is intended to present an important refection of the concepts. It also assists in understanding the main approaches used for solving mobility and presents a good background study and a start for the latter parts of the chapter where the actual design processes using the generic design model are presented.

---

[2] Hierarchical Mobile IP was also referred to as Regional Registration in [47][5] using the term often used for IPv4 specification of the protocol. .

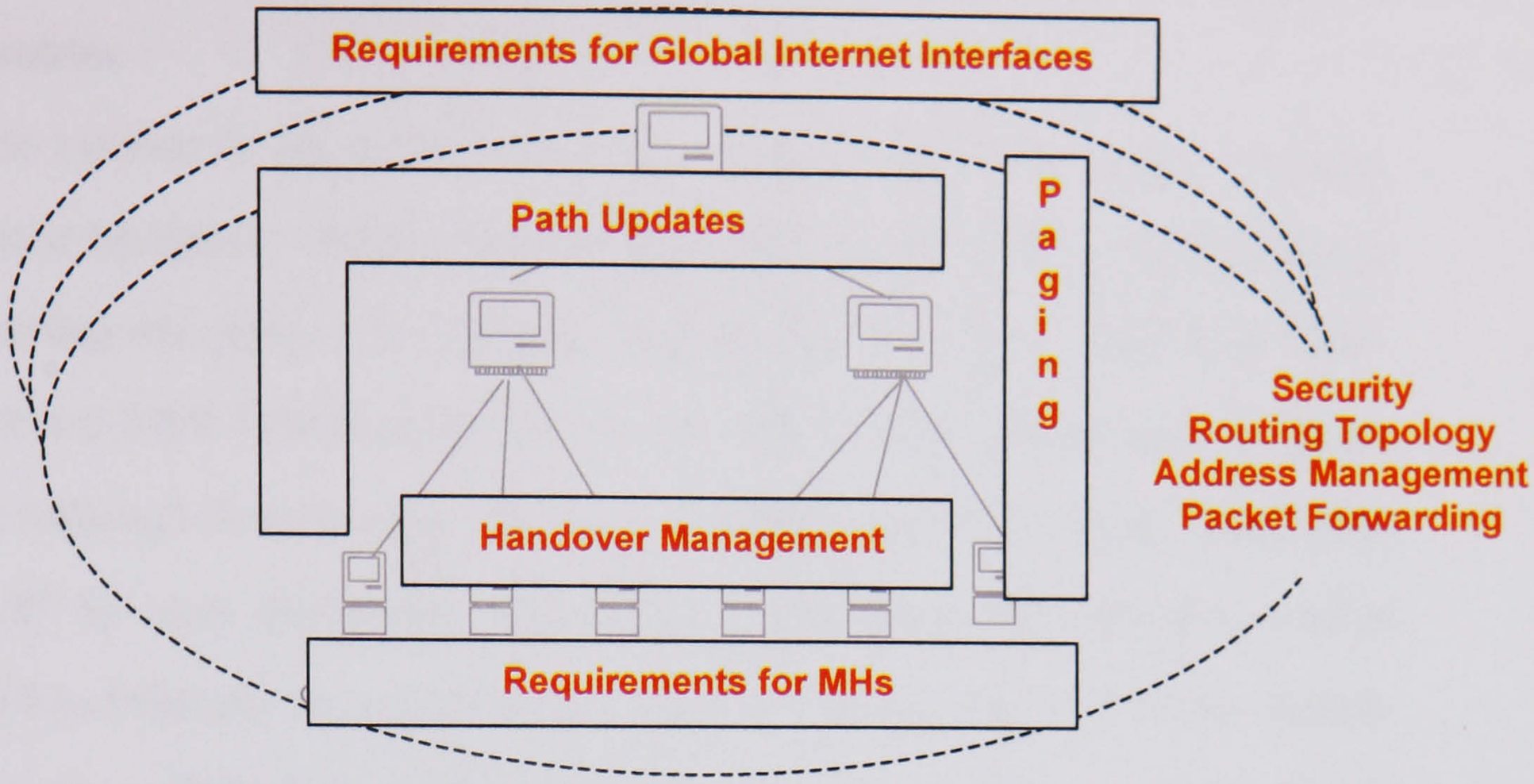


Figure 5.2. A generalised illustration of the Protocol Design Issues

5.2.2.1 Packet Forwarding (Routing)

Packet forwarding defines methods of packet delivery by mobility protocols to and from MHs. In the 'traditional' Internet, this is usually based on shortest path routing (for example, OSPF facilitated packet forwarding based on the destination address of the IP packet), where the aggregation of addresses means that routing can be prefix-based and carried out by the execution of routing protocols and rooting table look-ups. However, this must be modified in order to cope with host mobility as shown is Chapter 2 where the addresses allocated to MHs might not have a topological relevance, i.e. are not prefix routable to exact location of MH's current point-of-attachment. Typically, the packet forwarding solution in mobility protocols is based on **host routing entries**, with or without tunnelling. Host routing entries are a form of mobility-specific routing tables built by the rules set in mobility protocols and are not necessarily driven by the logic of the standard routing protocols. Host routing entries are a practical realisation of the distributed location information (i.e. LD) of the abstract mobility model shown in Chapter 2 while packet forwarding corresponds

to a form of distributed redirecting facilitated by the routing "pointers" existent in host routing entries.

The main contrast in the packet forwarding implementations is, on the one hand, Hierarchical Mobile IP, which extensively uses tunnels and default IP routing between the tunnelling end-points (i.e. Mobility Agents/FAs), and on the other hand MMP, HAWAII and MER-TORA, which rely on specially built routing methods for packet delivery (although these packets may be encapsulated from MH's HA). Hierarchical Mobile IP forwards downstream data within the domain using sequential tunnels between FAs (Mobility Agents). The host routes are implemented in Mobility Agents. This may be *inefficient* since handover latency depends on location and density of Mobility Agents (this property is already analysed in the handover performance analysis shown in the previous chapter) and because of the processing involved during the creation/termination of sequential tunnels and due to encapsulation overhead (although packet de-capsulation and encapsulation can be avoided by changing the IP addresses in the encapsulating header). This technique can be generalised as *cascaded tunnelling* where host routing entries are implemented in Mobility Agents (i.e. Foreign Agents) which use IP tunnels to route packets between them using the underlying IP routing. This is shown in Figure 5.3.

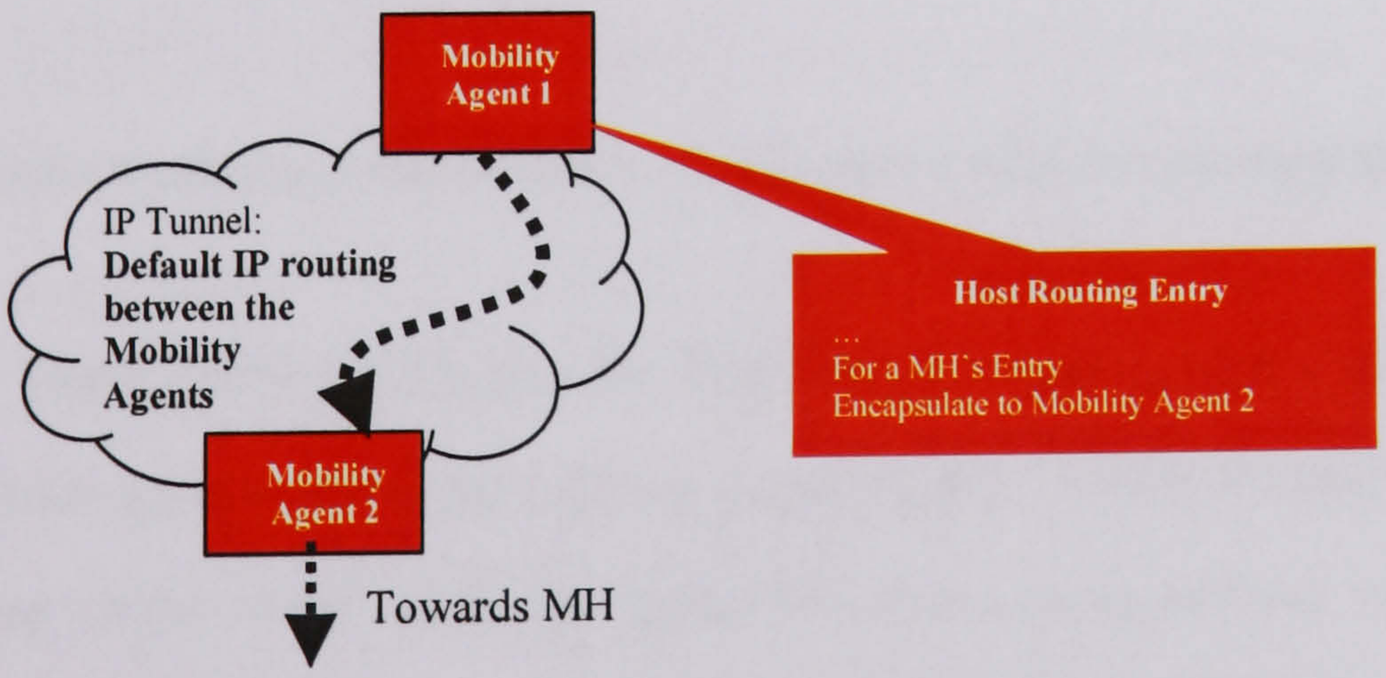


Figure 5.3. Conceptual representation of *cascaded tunnelling* packet forwarding technique

With MMP, packets are encapsulated by the ingress router (Gateway of the network) into multicast packets and are forwarded using CBT interface-based routing. The host

routes are implemented as CBT routing entries and additional FA-entries in the Gateway. In MMP, packets destined to another MH within the domain are sent up to the ingress router, which reverses them back to the target MH. In HAWAII, host routes are collections of per-host forwarding entries in the network. For Hierarchical Mobile IP, HAWAII and MMP, upstream packets can be forwarded with the same mechanisms that are defined for basic Mobile IP (for example, using reverse tunnelling back to the HA if the source addressing problems are to be avoided). Techniques for specific "overriding" of routing installations in the network used in MMP and HAWAII (and some other protocols such as Cellular IP) can be generalised as *host routes* where the latter uses it own methods of installing the host routing entries in routers while the latter uses the available mechanisms of CBT multicast routing. This is shown in Figure 5.4.

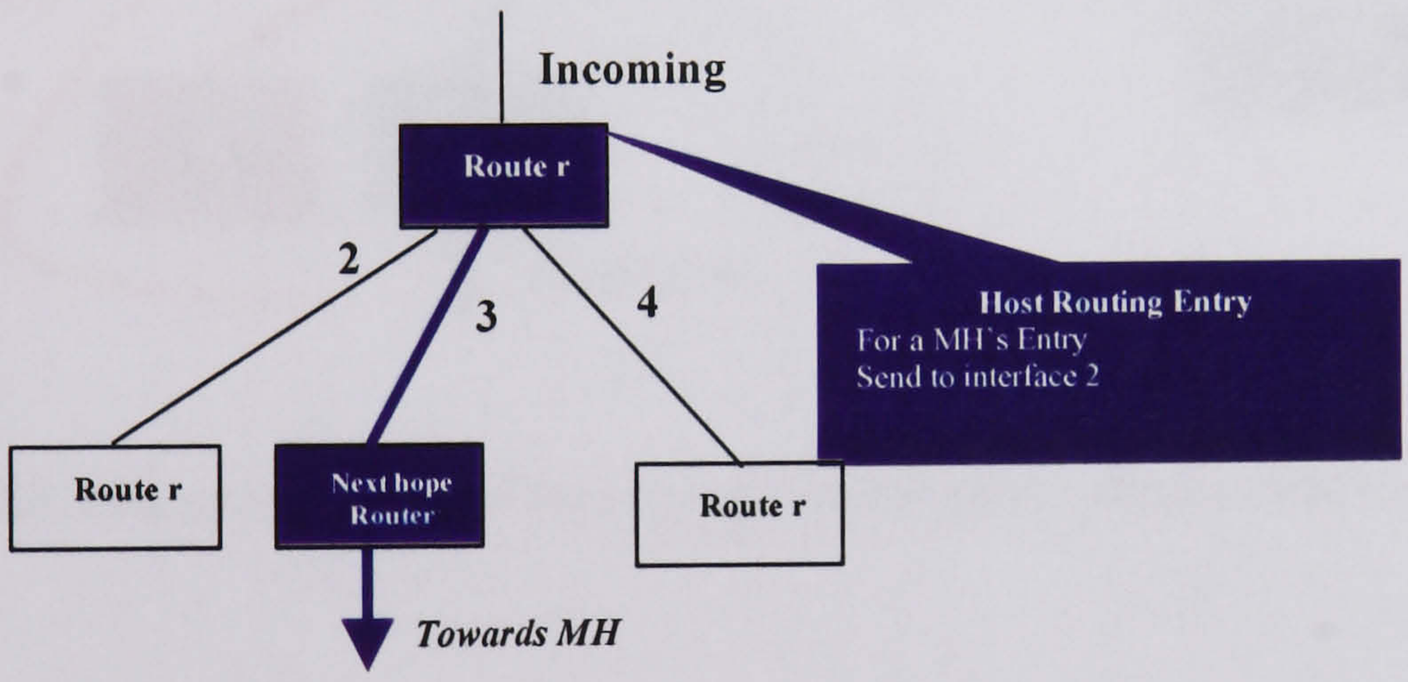


Figure 5.4. Conceptual representation of *host routes* packet forwarding technique

On the other hand, MER-TORA uses the TORA ad-hoc routing protocol for upstream and downstream packets from the address origin from/to which normal IP routing is used. Address origin is the initial BS hence MH receives an address with the same prefix as BS to which they initially attach. MER-TORA applies a *prefix-based routing* technique in two stages. Initial stage is *partial default prefix-based routing* to the address origin (i.e. in MER-TORA this is the initial BS) from whereon routing is

done by TORA ad-hoc routing protocol adopted for fixed networks constituting the *"hard-sate" prefix-based routing* to consequent points-of-attachments of a MH. The term "hard-state" explains that the TORA routing protocol used in MER-TORA overrides the underlying routing protocol and takes care of packet forwarding to current BS using its own packet delivery rules. Theoretically, *"hard-state" prefix-based routing* could be performed by the default routing protocol used in the network requiring updates of routing states for addresses given to MHs in order to route packets to their current BSs (more on these techniques is given in section 5.4). Essential *prefix-based routing* concept is shown in Figure 5.5.

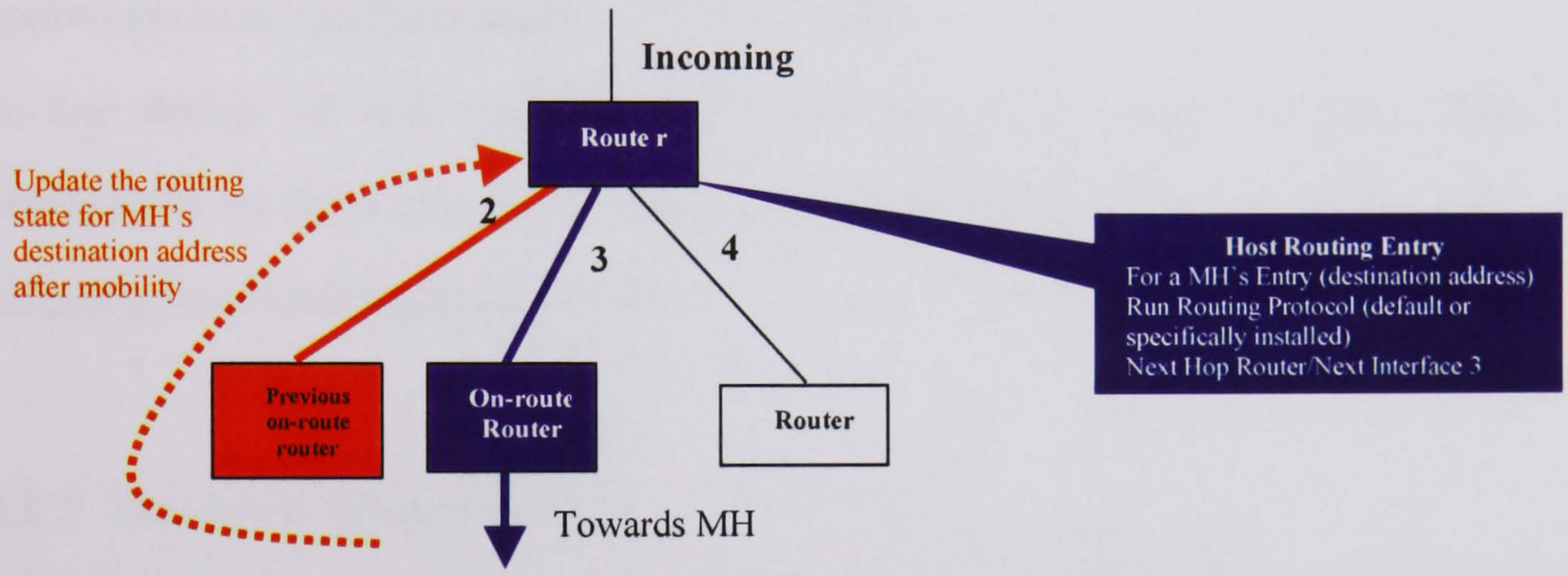


Figure 5.5. Conceptual representation of *prefix-based routing* packet forwarding technique

5.2.2.2 Path Updates

This Issue refers to the mechanism for installing information in the fixed network so that packets can be successfully forwarded to MHs at their new points-of-attachment. It typically consists of an intelligent transmission of *update* messages. In most mobility protocols this is solved by using the specific message format defined in the protocol or by using modified Mobile IP registration signalling model. Path Updates stand for the method of creating and maintaining the location directory (LD) described in the abstract mobility model.

This PDI also contains some interesting contrasts. HAWAII and MMP (CBT Join Requests and Join Acks) both use path updates for reconfiguring the tree of "soft state" entries, usually by only having to update/reconfigure the entries of the "cross over" router. This comes from their essential property of having an entirely host-specific routing. By contrast MER-TORA uses the path updates to update the "hard state" in the router, which means that when a MH changes its point-of-attachment MER-TORA creates a more host-specific state almost resembling the native prefix-based routing, which MER-TORA entries override. Finally, in Hierarchical Mobile IP, sending the Registration Requests and Replies between MHs and Mobility Agents reapplies the basic model of Mobile IP Path Updates.

The key feature of Path Updates PDI is the format of control messages, their destination in the fixed network and separation of messages that may be deployed for facilitating Handover Management PDI.

5.2.2.3 Handover Management

This Issue mostly considers the impact of handovers on MHs (whereas the previous two Issues took a more network-centric view). **Handover management refers to mechanisms for facilitating changes of points-of-attachment**. There is a variety of handover models distinguished by the methods in which the transfer of MHs to their new point-of-attachment are enabled and their impact on connectivity of MHs:

- The starting classification criterion can be the scope of handovers because that indicates the global impact of handovers with respect to the connectivity of MH's. A handover can cause a change of the MH's connectivity from one administrative domain to another. Thus, from the IP perspective, such handovers are called intra-domain and inter-domain handovers involving local (regional or micro) mobility mechanism or global (macro) mobility mechanisms respectively. This handover scenario naturally assumes IP transitions, because of the evident change of IP

domains. Inter-domain and intra-domain handovers are often termed differently. Usually, the different names represent identical handover processes and of the same handover types as specified in this paragraph. A slightly modified approach to handover taxonomy, also centred on the scope of handovers as the distinguishing criteria, separates handovers into vertical and horizontal handovers. While there is a direct analogy between inter-domain and vertical handovers (and intra-domain and horizontal handovers), vertical handovers often assume a heterogeneous environment where the change of IP domains involves an additional change of wireless access networks (for example an WaveLAN to UMTS handover). A change of IP domain as represented by the inter-domain handovers, even if they run different wireless access, can sometimes take a different form when the change also involves network administered access as represented by vertical handovers.

- Handover can be restricted to the link layer only, without a change of IP attaching interface inside the fixed network. This is a wireless access handover, or sometimes referred to as a link layer switch (for example a switch between Access Point Transceivers in HIPERLAN 2 wireless access technology [62]).
- Handovers can be further categorised according to the level of control given to MHs. Mobile or network controlled handovers are based on whether the MH or the BS (or another entity in the fixed network) decides on a handover, mobile or network assisted depending on which entity provides the information needed for the handover, forward or backward whether the new or old BS initiates the handover and planned or unplanned according to whether or not initial signalling takes place before the actual handover. Soft (facilitating the make-before-break scenario) or hard (break-before-make) handover depends on whether or not the MH simultaneously communicates with the old and the new BSs during a handover.

- Looking at the impact of handovers on the overall quality of the MH's connectivity: a smooth handover has a minimum packet loss, a fast handover has a minimum packet delay and a seamless handover is both smooth and fast.

Most of the IP mobility protocols mentioned handle handovers in a similar manner. The key operation is to perform the updating of the "cross-over" entity in the network (a router or a Mobility Agent) causing packets to flow to the new point-of-attachment. Variations exist due to the way in which Path Updates are implemented and how Packet Forwarding is performed. Some protocols (for example, MMP's advance registration procedure, Hierarchical Mobile IP, Cellular IP...) allow creation of simultaneous bindings to achieve bi-casting, i.e. duplication of packets from a "cross-over" entity to both the old and the new point-of-attachment thus assisting a smooth handover. Considering these handover procedures with respect to the Evaluation Framework, they are often extensions of Path Updates, or more appropriately re-runs of Path Updates for every handover. No additional handover support is required. Mobile IP (consequently Hierarchical Mobile IP can benefit from it), HAWAII and MER-TORA have proposed more complicated algorithms for handling handovers by allowing an exchange of messages between MHs and the old and the new points of attachment (i.e. BS). These efforts provide a foundation for further research in handover management, which is dealt in more detail in Chapter 6.

Both HAWAII and MER-TORA can optionally deliver, from the old to the new BS, packets that would otherwise be lost during handover. There are differences, however: in the Single Stream Forwarding sub-scheme HAWAII uses which it calls 'interface-based forwarding' meaning that the outgoing interface (on which to forward the packets to the new BS) is determined by both the IP address of the new BS and the incoming interface of the packet, whilst MER-TORA uses a temporary tunnel. However, in MER-TORA, if there is no tunnel when the link to the MH is lost (because handover may not be predicted), then a virtual link is constructed to the MH from the old BS. It retains this for some time in the hope that it will be notified of the

MH's new location. This virtual link should improve *robustness*, compared to the routing loops (or rather routing "knots" in cases of extreme non-optimal routing during handovers) that can transiently appear in some HAWAII sub-schemes because of the temporary redirection of packets from the old BS which travel "back and down" via the "cross over" router. Although this HAWAII model may cause a non-optimal routing and is partly overcome in the MER-TORA proposal, the overall conclusion is that more efficient algorithms are needed. The MER-TORA's handover support is more a necessity than a performance boosting feature due to the particular nature of the ad-hoc protocols, in particular TORA, which is adopted as the routing solution in the fixed network. In addition, both schemes are not offering generic handover solutions because of the dependence on specific Path Updates PDI Solutions.

5.2.2.4 Support for Idle Mobile Hosts/Paging

Paging reduces the frequency of refreshments/updates for an idle MH to achieve two goals: reduce the protocol overhead (signalling, route lookups and memory requirements) in the network and minimise the power consumption of MHs. A MH in the *active* mode is one that is sending and/or receiving packets via its network interfaces. Two further separate, but interrelated, modes are possible in order to optimise its mobility support. They can be coarsely characterised as follows:

- **Stand-by mode**: Its main goal is to save battery power in the MH, by allowing a MH (or subparts, e.g. its radio) to 'switch off' during a sleep period. Typically such a *stand-by MH* wakes up at well-defined times, during which the network can reach it. The stand-by mode enables link layer power management and so is only relevant to the particular wireless technology in use. Hence it is not considered further in this section.
- **Idle mode**: Its main goal is to reduce location update signalling over the air and in the network, by tracking the location of an *idle MH* less accurately than for an

active MN. The idea is to define a *Paging Area*, consisting of several BSs that correspond to some geographical area. Only when an idle MH moves into a new Paging Area (PA) does it have to send a location (paging) update to the network, whereas an active MH must send a handover/Path update message every time it connects to a new BS. Clearly, the network must also be able to re-activate the MH (for example, if a correspondent wants to communicate with it). The whole process of the idle mode support is often referred to as *Paging*.

Most mobility protocols, which provide paging mechanisms, are concerned mainly with paging in the network, i.e. how are the BSs in a paging area alerted that a MH in that paging area needs to be woken up.

The HAWAII proposal uses administratively scoped IP multicast [37] to distribute paging requests to BSs. This pushes paging to the edge of the IP network, which assists in *scalability* and *robustness*. A similar scheme is probably widely applicable in other IP mobility protocols. MMP naturally tracks MHs as they move, through the standard messages to join/prune from the multicast tree. It is suggested that the location management overhead may be decreased by reducing the refresh frequency of the CBT "soft state" mechanisms for idle hosts. A paging protocol has also been proposed for Hierarchical Mobile IP [63]. The protocol aims at independence of link layer technologies; the MH agrees a 'sleep pattern' with the network, which requires synchronised sending of Paging Agent Advertisements from FAs belonging to the same Paging Area.

5.2.2.5 Requirements for Mobile Hosts

An important factor in the design of every IP mobility protocol decision is to what extent MHs are required to participate in the establishment and updating of the routing structure that enables mobility. A reference example can be Mobile IP where a MH is

required to perform minimal operations: registering addresses, detecting movement and refreshing registrations.

HAWAII and MMP appear to have the *simplest* requirements on MHs, i.e. only MIP capability with extensions. In HAWAII the MH must be able to acquire a co-located CoA in a foreign network; MER-TORA suggests that a FA-CoA must be acquired. In Hierarchical Mobile IP the leaf FAs (lowest FAs in the hierarchy) support basic Mobile IP, which guarantees *compatibility* with dumb MHs.

This Protocol Design Issue is directly related to Handover Management PDI, which usually attempts to shift protocol functionality to MHs in order to facilitate a more efficient performance during handovers (more on their interrelations is given in section 5.4).

5.2.2.6 Requirements for Global Internet Interface

This issue defines the functionality in the ingress/egress router of the IP network usually referred to as a Gateway. A Gateway is the transition point between the global and regional mobility or, as referred to in earlier parts of this document, the separation point between the scoped network, where the regional mobility protocols are deployed (for example, the *micro mobility* section of MMP), and the rest of the Internet. Gateways can include functions such as interworking between the regional (micro) and global (macro) mobility, mapping of addresses, tunnel management, central control of mobility protocol mechanisms…

A common objective in many IP mobility protocols is to minimise changes to the standard IP protocols. All schemes seem to make some additional requirements on HA operation (limiting *applicability*); for instance, the Gateway in MMP (similar in HAWAII) functions as with respect to the home network (HA) of the MH. In such a setup, the Gateway may additionally have a specific security association with the HA and also the MH. MER-TORA can have several network gateways (aiding *robustness*

and *scalability*), whereas the others appear to be able only to have one. This comes from the assumption that the administrative gateway of a network is the ingress/egress point of the micro mobility (e.g. Gateway in MMP and top-level Mobility Agents in some *Proxy Agents Architectures*). This can be avoided by placing the mobility gateway(s) at arbitrary points inside the network bearing in mind the logical area controlled by the mobility gateway.

5.2.2.7 Address Management

A MH typically has to be provided with an IP address in a visited network. The way this is done can have an important impact on, for example, handover performance, scalability (because in IPv4 for example unicast addresses are becoming a scarce resource), and deployability (in some corporate IP networks private home addresses may need to be supported to overcome firewalls). **This Issue relates to overall implications of different types of addresses used by MHs during mobility.**

Address management is a key issue and a significant contrast between the protocols. With HAWAII, MMP and MER-TORA a MH keeps its IP address (unicast or multicast *care-of-address*) throughout the lifetime of the session, at least while it is in the same domain. This would (for example) ease the *applicability* of RSVP-based QoS support [64][3], which uses the IP address for identification of the QoS flow mappings. By contrast, in Hierarchical Mobile IP *care-of-address* changes at each handover. HAWAII requires that in a foreign network a MH acquires a publicly routable co-located *care-of-address*. Given the scarcity of public IPv4 addresses, this is a major drawback from the point of view of *scalability*. Also, because the *care-of-address* must be unique within a domain, a co-ordinated address allocation mechanism must be available (a similar method for address allocation method is also proposed for MMP's multicast *care-of-address*, as suggested in section 3.4.1). Hierarchical Mobile IP can also use a co-located *care-of-address,* and then similar comments would apply.

But, instead it could use a *FA-care-of-address* and then IPv4 address exhaustion is not a problem. Within the domain, private *care-of-addresses* can be used since they are not visible outside the domain. In MER-TORA, a MH is allocated an IP address by the BS where it starts a 'session', from the IP address block that it owned by the BS. The pros of such a procedure are: fully prefix-based routing until MH moves minimise host-specific routing overhead and allows for a consistent and *simple* address allocation across the domain since each AR owns its own address block. The cons are: more addresses are probably needed than for a IP mobility scheme with flat addressing across the domain, and more frequent address de-allocation is required (for *scalability* the IP address should be returned as soon as possible, for example, at the end of an active session and not just when the MH powers down). If the number of MHs is large and their sessions short, then clearly a good, scalable DHCP implementation is needed. In MMP, the MH acquires a multicast *care-of-address*, which requires intelligent distribution of addresses to avoid shortages of IPv4 multicast addresses and overlaps with existent pure-multicast sessions. However, MMP uses a scalable method because of the recycling of local multicast addresses and the transparency of these are chosen for the administratively scoped "multicast address pool".

The scalability problems are less present in the IPv6 adaptation of the IP mobility protocol, mostly due to significantly larger address space.

5.2.2.8 Routing Topology

This refers to a general static view of the IP network elements and their potential impact on mobility protocol operations; whilst the Issues mentioned previously more or less cover dynamic protocol operations. It refers to the arrangement of the elements (for example, whether they must form a hierarchical (tree) topology) and their required capabilities (for example, whether they can act as normal IP routers

and/or IP capable BSs). The routing topology has implications on the scalability and robustness of the system, i.e. robustness may be a problem if the IP network hinges on a single gateway node. This Issue also relates to the reaction to any failure of links or routers.

Clearly, the relevant routing protocol capability needs to be *deployed* using adequate elements in the network. The effort is probably greatest for MER-TORA, because standard unicast routing is replaced by TORA. However, MER-TORA authors argue that it will give *scalability* advantages. *Robustness* is achieved in all "soft state"-based protocol which is mostly a feature of all *Localised enhanced-routing schemes*. However, variations exist: HAWAII relies on standard routing protocols for detecting failures; by integrating HAWAII with a routing daemon, a change in the default route can trigger soft-state refreshes to HAWAII paths. MMP relies on the "soft state" mechanism of CBT to detect the reachability of upstream neighbours. *Proxy Agent Architectures*, such as Hierarchical Mobile IP, rely on standard protocol recovery mechanisms to adapt to changes and failures. Hierarchical Mobile IP uses centrally rooted tunnelling trees over arbitrary topologies whilst the setup of other protocols matches the underlying topology which can be tree or mesh.

5.2.2.9 Security

Unlike fixed Internet communications, mobility, and wireless access in particular, introduce intricate security issues: the user's access to a visited network needs to be authorised and any requests for alterations of protocol functions (routing entries...) have to be authenticated; the user's privacy should be preserved; the network's topology could be hidden from MHs; interworking with end-to-end IPSec should be allowed... The majority of IP mobility protocols include frameworks for realisation of security features by typically assuming interworking with security protocols already available. This comes from the fact that

most of the security issues are generally related to the setup of the network and MHs, rather than the particular features of IP mobility protocols. Hence, they can be collectively analysed.

As mentioned in MMP security considerations in section 3.4.3.6, most of the security issues for IP mobility protocol relate to securing transfers of protocol control messages and critical functions performed by the visited network. Initially, a user has to be authorised to access and use the network. This can be performed by the available authorisation mechanisms such as AAA [71] (AAA can also negotiate a session key between the network and the mobile host). IP mobility protocols generally aim to use authenticated protocol control messages, which are mostly related to the mechanisms of Path Updates, Handover Management and Paging. Authentication of data packets is generally not considered due to the associated cost of the transport and the use of available end-to-end security mechanisms (IPSec). The most critical part of the security procedures occurs during handovers because of the requirement for fast handovers and the need to have the session key (used by the MH and the routers/Agents containing the routing entries) available at new points-of-attachment (BS/Agents) and the rest of the visited network, previously uninvolved in the communication with the MH. The standard authentication procedures often require signalling support for transferring the information on a MH's session keys and are generally slow, hence only suitable for Global mobility, i.e. inter-domain handovers.

The work is under way to enhance the AAA protocols to enable fast session key management for micro mobility scenarios. Mobility protocols can assist these operations as explained in Chapter 6 where a context transfer can be performed during handovers to convey security data for MHs. Some previous solutions for providing dynamic availability of session keys throughout the visited network (Cellular IP) propose algorithms for fast calculation of session keys at new points-of-attachment and the rest of the mobility infrastructure. The essence of the algorithm is to calculate the session key in the visited network based on the IP address of the host, the network

key and a random number (for replay protection). This is returned to the MH during the initial login and used by all nodes involved in the network. If no link layer security is deployed, the network layer key can be used for encryption of data sent over the wireless link.

HAWAII uses DHCP for initial address allocation and expects execution of authorisation and authentication procedures at the same time (AAA can assist such a procedure). The same principle could apply to other mobility protocols, which obtain *care-of-addresses* in a collocated fashion, i.e. for the duration of the connection in the network (MMP).

Additionally, for mobility setups, which use foreign network functionalities to impersonate Mobile IP functions on behalf of MHs (HAWAI and MMP), there needs to be a verification and key-management infrastructure to distribute temporary session keys to the MH, foreign network and HA. In MMP, this requires the Gateway/Core to have a security association with the HA. Additionally, MHs need to trust the Registration Replies generated by the Gateway/Core and the HA needs to trust the decapsulation/encapsulation performed at the Gateway/Core when functioning as a FA on behalf of the attaching MH.

Besides the Mobile IP-based control, MMP has a slightly relaxed security problem regarding the control messages and the routing states that they may create in the network. This is because the CBT part of MMP, is entirely managed by the network, i.e. the CBT messages (Join Requests…) are generated by BSs (or other routers in the network) and not by MHs. Hence they can be easily authenticated since they are only used inside the network and are generated by the network entities. This is a highly beneficial feature of MMP, because it separates the (surrogate) Path Updates from the registration messages (Mobile IP-based).

Security has received limited consideration in other mobility protocols. In general, it is suggested that existing mechanisms can be used; for example, Hierarchical Mobile IP mostly refers to the existing Mobile IP related security infrastructure.

	Hierarchical Mobile IP	Multicast for Mobility Protocol	HAWAII	MER-TORA
Packet forwarding (downstream)	Cascaded tunnels	*Host routes:* multicast forwarding of encapsulated packets	*Host routes:* specific HAWAII techniques for end-to-end encapsulated packets	*Default Prefix-based* route to BS; *"Hard-state"* to new BSs
Path updates	Mobile IP + H.MIP extensions	CBT control messages + MMP Instruct + Mobile IP signalling for registrations	HAWAII Path Updates + Mobile IP Signalling	Path Update message from old-BS to new-BS for installing hard state, host-specific routes
Handover management	Mobile IP, Route Optimisation	Mobile IP, Multicast join, Advance registration, Simultaneous bindings (network managed)	Forwarding/Non-Forwarding schemes (planned/unplanned handovers)	Localised at the edge of the network; inter BS tunnelling
Support for idle MHs	No, separate proposal [63]	Reduced signalling in wired network	Paging using IP multicast, separate proposal [37]	No
Requirements for MHs (in addition to basic MIP support)	Extended Mobile IP features	Adopted Mobile IP features, multicast CoA	Adopted Mobile IP, Route Optimisation	TORA, address acquisition, tunnel initiation, address return
Requirements for global internet interfaces	Dependent on the placement of Top-level Mobility Agent/FA. In case collocated with the gateway packets need to pass via the specific gateway	Core/Gateway collocation typically assumed.	HAWAII Gateway typically assumed to be collocated with the network gateway.	Gateway's functionality not critical nor the number of gateways in the network
Address management	Typically FA-CoA (link local stateless). Could adopt Co-located CoA (bypasses the domain hierarchy)	MH retains a multicast IP address within the domain. Ingress router seen as FA.	Static Co-located CoA in Foreign domain, Home Address in home domain	Initial BS (address origin) allocates an IP address from set it 'owns'. De-allocated at session end.
Routing topology	Static configuration of enhanced Mobile IP FAs in a tree structure	All nodes must support CBT IP multicast (sparse mode). Route created using shortest (assumed) path CBT message traversal to Gateway	HAWAII-aware routers; standard Routing protocols keep the default route up to date.	All routers implement TORA ad-hoc (finds the shortest route available)
Security	Mobile IP (Route Optimisation) security features + supports authentication protocol	Assumes security association between MH, Gateway and HA. Adopts control messages encryption mechanisms and Mobile IP security features	Assumes security association between MH, FA(ingress router) and HA. Adopts control messages encryption mechanisms and Mobile IP security features	Largely assumes adaptation of existing mechanisms for authentication

Table 5.1. A summary of how the exemplar protocols tackle each Protocol Design Issues

5.3 Principles of the Generic Mobility Design Model

The previous section contains description of the Protocol Design Issues (PDI) and their Solutions in example mobility protocols chosen from each category of mobility

protocols defined in section 2.3.3 including MMP. Introduction of PDIs as segments of overall mobility functionality in consideration of mobility protocols offers several conclusions for further analysis and design of mobility protocols:

a) The first conclusion of the overview of IP mobility and application of the Evaluation Framework is that there is a conceptual and functional split between the features of IP mobility protocols identified by the PDIs (see Figure 5.2) and their Solutions. The extent and properties of the relations between the PDI Solutions is analysed in the latter parts of this section.

b) Introduction of PDIs provides foundation for comprehensive analysis and evaluation of all features of mobility protocols. The advantages and disadvantages of a mobility protocol can be more evident when analysis is conducted considering each PDI and specific set of evaluation criteria. In addition, mobility protocols can be compared along the "axis" of each PDI, which can offer a rigorous and revealing extent and types of their differences. Section 5.3 offers an example comparison of MMP, Hierarchical Mobile IP, HAWAII and MER-TORA.

c) Application of PDIs in representing collective mobility functionality provides a starting point for modular design of IP mobility solutions. From a high-level design perspective, some functionality of mobility protocols can be separated. Hence, a protocol can be designed by choosing the appropriate mobility modules, i.e. **primary PDI(s)**, as main design targets and constructing the rest of the protocol from the remaining PDIs provided this is consistent with the primary choice. This section contains detailed description of theory of using each PDI as the primary design choice and resulting relations to the rest of the PDI set. This can then be used as a conceptual tool for modular design of mobility protocols. Finally, regarding the actual features of primary or remaining PDIs these can be separately developed by constructing novel PDI Solutions (i.e. mobility features) or adopted from the available PDIs already present in existing mobility protocols. This is dependent on the desired functionality of the chosen PDIs and their potential

existence and functionality in the available solutions. This point follows assumption that evaluation of existing mobility protocols considering relevant deployment scenarios does not result in undisputed and sufficient selection of exiting mobility protocol(s).

The last point above is used as a guideline for the remaining description of the mobility design model and application of modularity as the main principle of the model. The main issue for realisation of the model is selection of primary PDI(s) and design decisions that enforce such selection. Primary PDI(s) can be chosen in different ways. One way would be to apply appropriate evaluation criteria and select PDIs that dominantly influence assessment of validity of protocol's performance with respect to the chosen evaluation criteria. Another way is to use particular design principles (for example taken from operator's preferences or design principles of network development as applied in the BRIAN project) for choosing the appropriate PDI(s) or for emphasising particular evaluation criteria. In any case, this initial step is subjective and can be implementation-driven, as it would depend on particular preferences for development of mobility solution. This separation between evaluation criteria and design principles as instruments for selection of primary PDI(s) is intentional and describes differences in their roles in processes related to the mobility model. In fact, evaluation criteria and design principles can often be the same. For example: any of the *efficiency* evaluation criteria (see section 5.2.1) can be naturally considered as a design principle (the differences may exist in the level of importance of some of the elements of the evaluation criteria from the set of *efficiency* criteria). In addition, evaluation criteria *ability to support dumb MHs that are Mobile IP compliant* (see section 5.2.1) can also be an explicit design principle. One the other hand, one design principle can be preservation of addresses for MHs and as such is not recognised as evaluation criteria (section 5.2.1) since it may depend on subjective design requirements and its benefits are difficult to identify from a general stance (hence it is not recognised as an evaluation criteria). The intention here is not on identifying the

clear differences between the evaluation criteria and design principle but on their roles in analytical and design tools presented in this chapter. Evaluation criteria is used for validating the basic quality of a mobility protocol or some of its PDI Solutions while design principles are distinct rules posed on mobility solutions which are often implementation-driven and are involved in the initial development stages (note: evaluation criteria can also be used in the initial stage of selection of primary PDI(s) as described above or for the general analysis and evaluation of already developed mobility solutions). One example approach in selecting primary PDIs is shown in section 5.4 in the design of mobility solutions in the BRAIN project. The remainder of this section discusses the extent and manner in which the modularity can be accepted in the design processes of mobility protocols. From an abstract perspective **modularity** offers the following advantages:

- Easier design: breaking up the problem into smaller pieces allows for more effective focusing on specific design issues (i.e. PDIs) and easier interpreting of the required functionality. Additionally, it provides a good starting point for evaluating and reusing the existing features of mobility protocols.
- Easier evolution of the design: dividing the problem along clear functional boundaries between the modules allows for separate evolution of the modules.
- Easier deployment: there is a potential for allowing standardisation of specific modules (e.g. Handover Management PDI Solution), whilst an operator may choose appropriate remaining modules to suit the particular deployment scenarios and requirements. This is achievable provided the modules (PDI Solutions) can be functionally separable as the remainder of the section examines.
- Operational stability: by deploying separate modules which compose the whole mobility protocol, failure of one module can be easily spotted and recovered by a "fall back" procedure to the operational modules. This point only applies to specific modules of the design, which can achieve a substantial level of functional

dependence (e.g. failure of the paging modules can spark a "fall back" to regular Path Update procedures for active terminal, failure of the Handover Management can be temporarily fixed by Path Updates…).

At the same time modularity incurs a certain level of complexity in interpreting and designing mobility protocols and may allow for expansion of the design proposals (for separate modules, i.e. PDI Solutions) rather than their convergence.

The next step in the explanation of this design model is the assessment of design significance, functional independence and impact of each PDI and its potential Solution. In order words, **the aim is to show how each PDI can be used as a driver of the whole design and the extent and consequences of such module-centric design approach, i.e. use of each PDI as the primary PDI and consequences of the selection on functionality of the remaining PDI(s).**

The following paragraphs include a short explanation on how each identified PDI can be used as the design driver for the overall mobility protocol and functional implications of their ultimate Solutions. **It should be noted that the primary guideline and objective of the following description is in achieving maximum level of independence of each PDI Solution.** This can be interpreted as the default method for understanding interrelations between PDIs and their Solutions, which can then be used for more specific development of IP mobility protocols where specific PDIs are "highlighted" due to the specific design and deployment requirements. One such example is shown in section 5.4.

Packet Forwarding (Routing)

The choice is between the Packet Forwarding techniques defined in section 5.2.2.1:

a) *Host routes* (e.g. MMP, HAWAII, Cellular IP…). Network is "infected" with specific *host routes* in the form of "pointers" to next hops downstream to MHs. Such scenarios make the address given to MHs inside the micro-mobility domain topologically irrelevant since the *host routes* in the network enforce routing (hence

this influences Address Management PDI). Examples of "prefix-less" addresses include multicast CoA in the case of MMP and permanent CCoA in HAWAII, which does not change during handovers, that is, changes of subnets inside the domain. *Host routes* are installed from the micro-mobility Gateways to which routing is facilitated by a topologically correct address in the outside Internet (i.e. via Mobile IP). In MMP, HA keeps the address of the Gateway, in HAWAII CCoA is prefix-routable in the Internet to the ingress point of the network, which is the Gateway. Hence, this packet forwarding technique assumes an explicit entry into the micro-mobility domain from where *host routes* are utilised. This imposes direct relation with Requirements for Global Internet Interfaces PDI. One example of the concept of *host routes* technique and its relation to some other PDIs is shown in Figure 5.6.

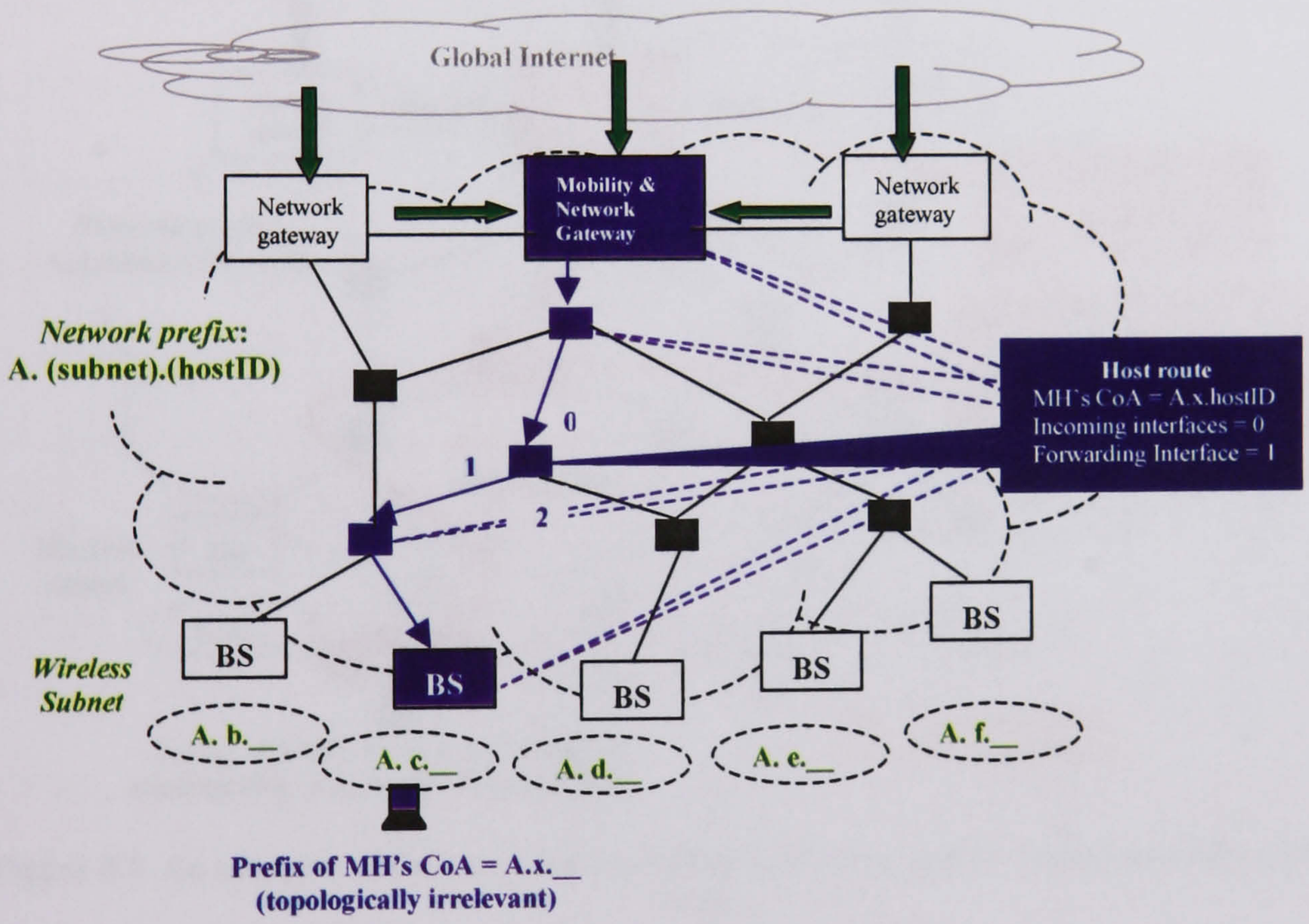


Figure 5.6. An example of the *host routes* technique and its relations to other relevant PDIs

b) Cascaded tunnelling[3] (e.g. Hierarchical Mobile IP. Tunnels are built between the Mobility Agents "down" to MH (IPv6) or its link-local Mobility Agent (a case in

[3] Note: Protocols such as HAWAII and MMP deploys distributed host routes in router inside the micro mobility domain. Although these packets may be tunnelled (end-to-end and from the Gateway the

IPv4). Routing is the default intra-domain prefix-based routing based on the destination address of the outer IP header. MH obtains a topologically correct address, either in a collocated fashion or from the link-local Mobility Agent. Address change is mandatory during handovers since the tunnelling between Mobility Agent(s) and MH is based on standard IP routing using the topologically correct destination IP addresses (topologically correct signifies that the destination of packets is found using the underlying routing protocol since the destination address has the same prefix as the destination router/BS). One example of the concept of *cascaded tunnelling* technique and its relation to some other PDIs is shown in Figure 5.7 (Mobile Agent and Network Gateway are collocated relating to Requirements for Global Internet Interfaces PDI).

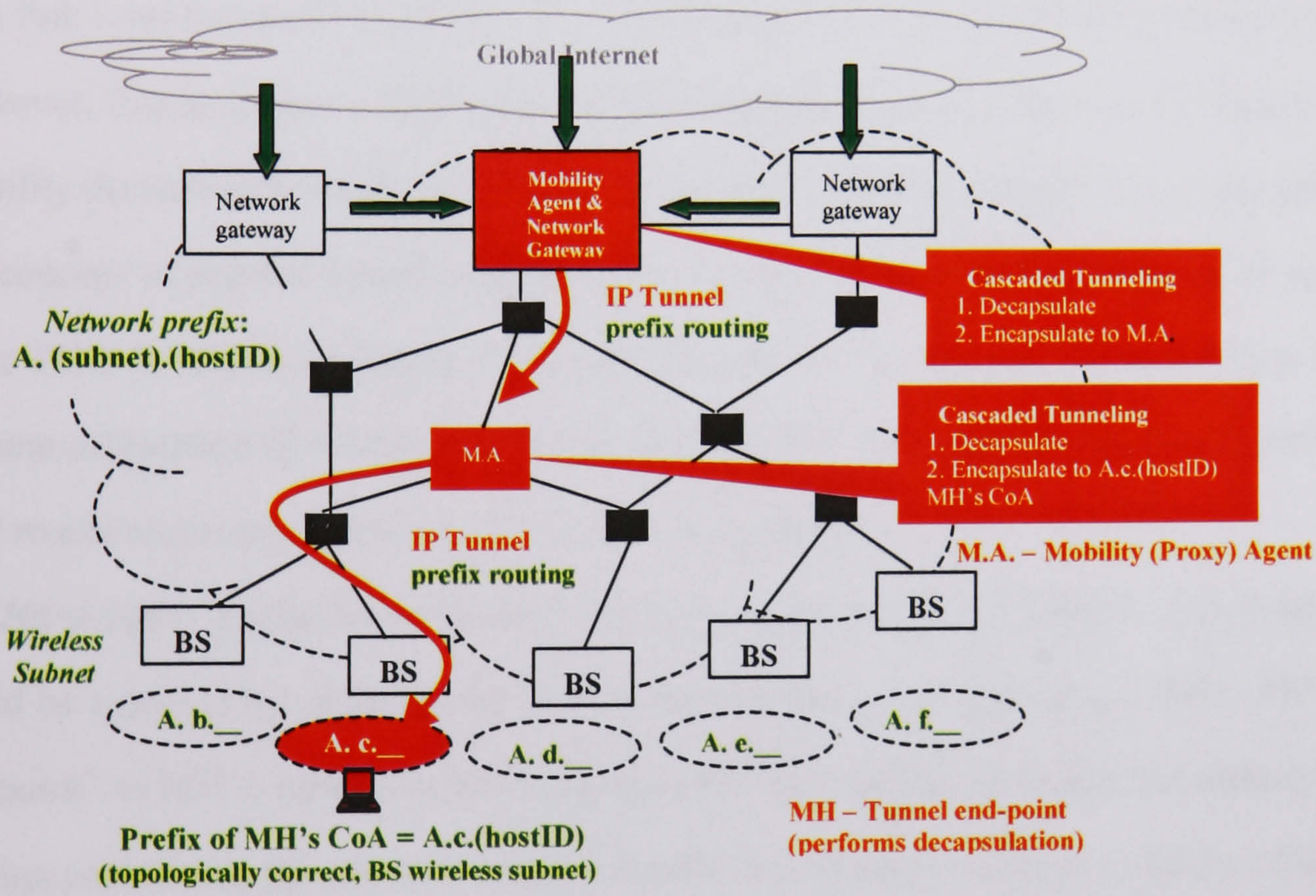


Figure 5.7. An example of the *cascaded tunnelling* technique and its relations to other relevant PDIs

c) Partial default prefix-based routing (e.g. MER-TORA to address origin, i.e. origin BS). Upon the initial registration/login MH obtains an address (CoA) from an "**address origin**", which can be any router in the network to which the address is

---

tunnelling techniques used in HMIP (*Proxy Agents Architectures*) distinguishes *cascaded tunnelling* because of the explicit encapsulation/decapsulation between Mobility Agents and the associated

prefix routable inside and outside the domain. In other words, the address has the same subnet prefix as the address origin. In MER-TORA address origin is the initial BS to which the MH connects when it initiates its connectivity in the network. This solution is partial in the sense that routing is only achieved up to the address origin. During the consequent mobility of MH to other BSs in the network one of the three other techniques need to be used to route packets to MH's current point-of-attachment which changes as a consequence of mobility, i.e. handovers to other BSs in the network. This technique can be observed as a hybrid case of the previous two since their mobility setups assume an abstract address origin to which the packets are prefix routable. This point could be the Gateway (e.g. HAWAII) or Mobility Agent (e.g. with link level connectivity to MH) in *host routes* or *cascaded tunnelling* respectively. However, this technique additionally facilitates prefix-routable CoAs inside the micro-mobility domain and offers arbitrary locations of the address origins. One example of the concept of *partial default prefix-based routing* technique and its relation to some other PDIs is shown in Figure 5.8 (Note: this packet forwarding technique does not assume collocation of Mobility Gateway and Network Gateway as usually assumed in *host routes* and some scenarios of *cascaded tunnelling)*.

d) "Hard-state" prefix-based routing: *Host routes* and *cascaded tunnelling* techniques could be replaced by updating the intra-domain routing protocols (e.g. OSPF, RIP…) to "point" to MH's current point-of-attachment. This would mean that the underlying routing protocol in the network has to be updated with the new route to MH's address at its current point-of-attachment with obvious scalability risks and delays during convergences of routing states. In such a case the topologically irrelevant ("prefix-less") CoA would artificiality acquire topological relevance or the prefix for a topologically correct address could be altered during mobility. Alternatively an overriding routing protocol could be deployed "on top" of the default routing such as TORA ad-hoc routing protocol used in the case of MER-TORA which combines

---

forwarding.

partial default prefix-based routing with TORA "hard state" routing from the address origin. One example of the concept of *"hard-state" prefix-based routing* technique and its relation to some other PDIs is shown in Figure 5.9.

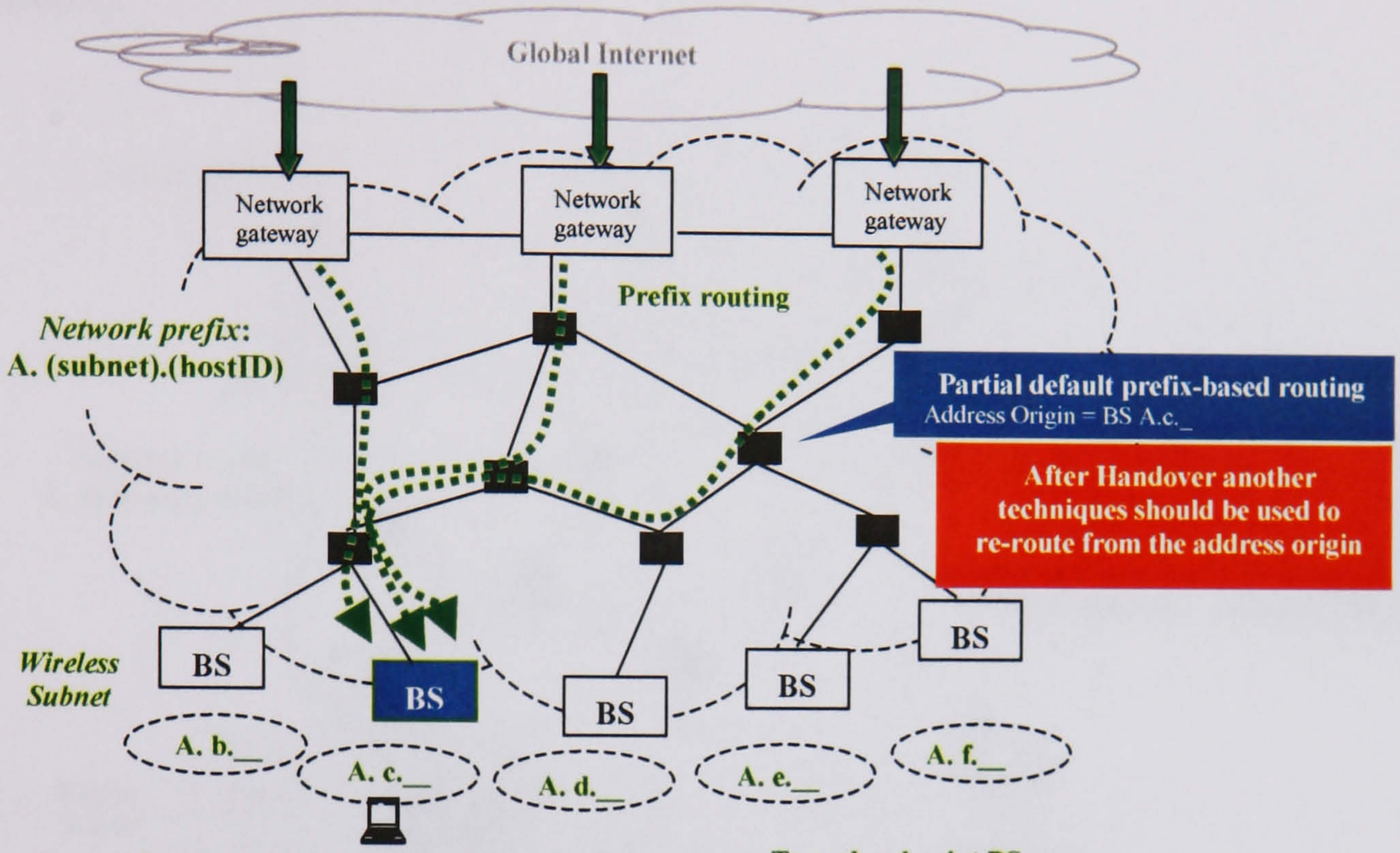


Figure 5.8. An example of the *partial default prefix-based routing* technique and its relations to other relevant PDIs

As the following text on remaining PDIs verify, when Packet Forwarding is influencing the design process of a mobility protocol this allow for considerable design freedom in choosing the Solutions for Path Updates (to the extent of timing and format of Path Updates not the destination of the messages), Handover Management, Support for Idle Hosts, Requirements for MHs and Security .

Packet Forwarding techniques as explicit design choices, thus design drivers, can be analysed in various quantitative and subjective qualitative manner. The quantitative analysis may weigh the overhead of using host routes, tunnels and routing installations as well as their impact on handover performance (results in Chapter 4 can be used as a start using *host routes* for MMP and *cascaded tunnelling* for Hierarchical Mobile IP).

Other criteria may include utilisation of prefix-based routing because of the reduced mobility related functionality or any other relevant accompanying support, limiting Mobility Agents, limiting mobility-specific software in routers, deployment scenarios…

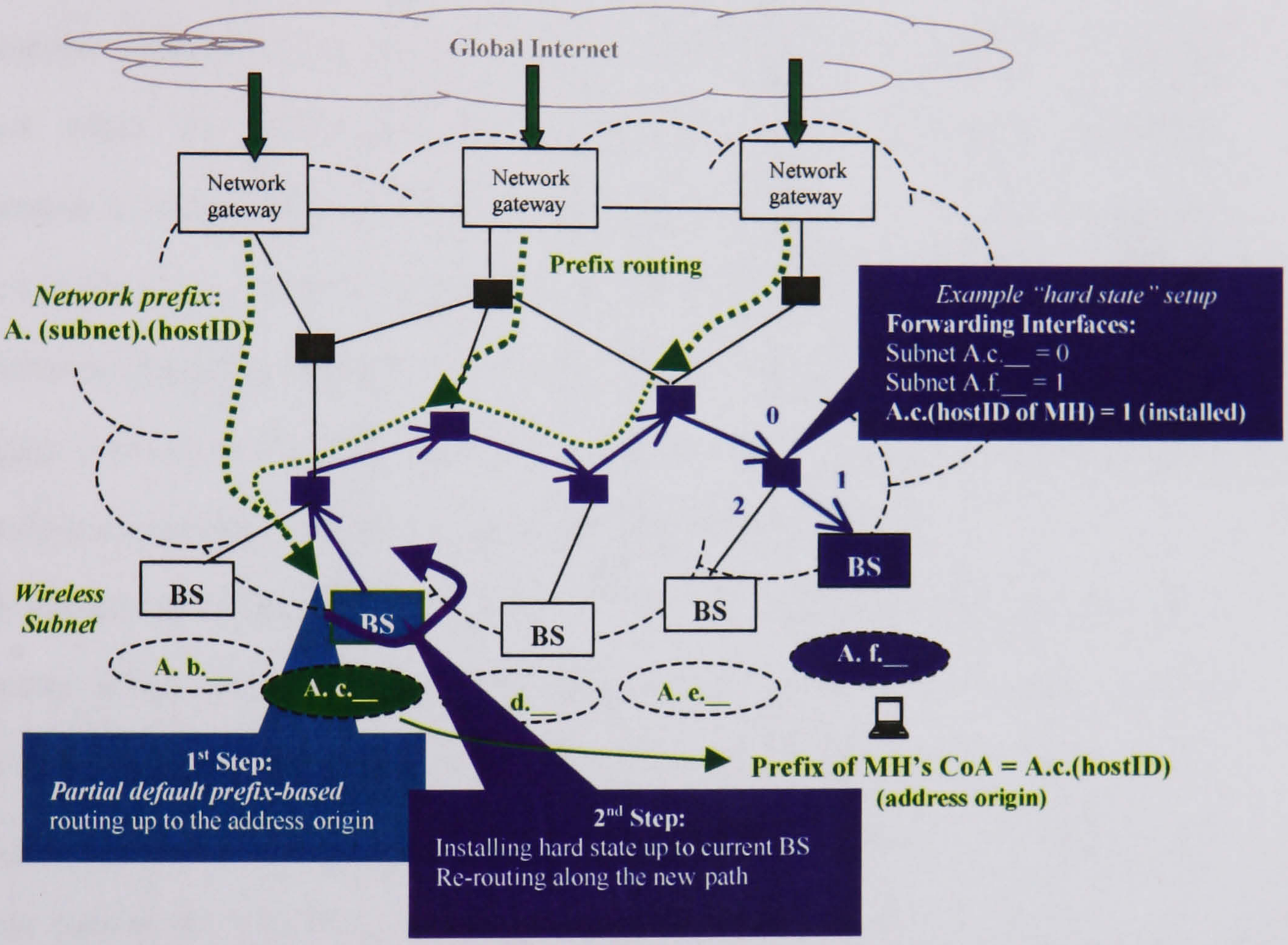


Figure 5.9. An example of the *"hard-state" prefix-based routing* technique and its relations to other relevant PDIs

On the other hand, Packet Forwarding techniques have direct impact on the Solutions for Address Management, Requirement for Global Internet Interface and some aspects of Routing Topology. If any of these were used as the design drivers, Packet Forwarding would be reduced to one or more of the available techniques, which conform to the specific design choices for those three related PDI Solutions. These particular relations are discussed below in the explanation of the three PDI Solutions.

Path Updates

Following the principle of maximum independence of PDI Solutions the first consideration is that the mechanisms of Path Updates are separable from Handover Management and Paging. Further assuming that Hanover Management Solution takes care of the packet losses during handovers, importance of Path Updates with respect to the *handover latency*, thus packet losses, is mitigated. Path Updates should be triggered when the BS-transfer (i.e. handover) facilitated by the Handover Management is certain. Path Updates can be triggered by MH or BS. MH is inevitably involved in the stages of the Handover Management which is decoupled from the rest of mobility so it may be beneficial to "hide" the process of Path Updates from MHs and trigger them from BS. This may also benefit security mechanisms and it follows the principle of maximum independence of each PDI Solution.

In such scenarios design of Path Updates is not rigid and can be left to operator's preferences when choosing a particular messaging system for updating relevant entities in the network (this being dependent on the Packet Forwarding setup and the resulting relations with other PDIs but only to the extent of where the message is sent to not the format and triggering of it).

Handover Management

As with Path Updates, Handover Management can stand as an independent sub-protocol inside the overall mobility protocol. As such, its design can be based on particular preferences briefly analysed in 5.2.2.3 such as the choice of control and assistance in either networks or MHs, dependency/independence on wireless scenarios… One such input of particular preferences is shown in the complete design of Handover Management Solution shown in Chapter 6.

Additionally, there should be no direct relation and interdependence between the Address Management and any stage of the execution of Handover Management. This also applies for most of the other PDIs and their Solutions but is specifically

mentioned for Address Management due to the primary objective of maximum independence of PDI Solutions and the novelty of such an approach (this is further elaborated in Chapter 6).

Support for Idle Hosts/Paging

Paging mechanism can be realised in an almost independent manner considering the Path Updates and Handover Management Solutions for active hosts (as their feature are assumed to be applicable to). The way this is achieved is a particular design choice and techniques may vary such as the "explicit" paging where a specific message is used to "wake up" MHs (e.g. HAWAII, MMP) or the "implicit" paging where data packets are sent along the paging route to reinitiate the active mode support (e.g. Cellular IP). Both solutions assume independent control messages for updating and creating paging entries in static and mobile scenarios. Another important consideration in the "explicit" and "implicit" paging solutions is that the latter can be realised through a limited (not all the way to the serving BS of the MH) chain of paging host routes which would most likely coincide with a similar (*host routes*) Solution/technique for packet forwarding. As such dependency (or rather assumed coexistence) is not recommended in the pursuit for maximum independence of PDI Solutions, "explicit" paging stands as the most promising solution as it is allows for independence from the Packet Forwarding setup. Paging solutions should be triggered after the expiration of the active states either by the MH (through a transmission of paging control message) or automatically by the network. Reverselly, paging should be deactivated by MHs assuming "implicit" paging is used in which case a control packet can be sent to the MH to reactivate the active mode support in case the activation is not requested by the MH but via the arrival of packets sent to it.

Requirements for MHs

As mentioned in section 5.2.2.5 the basic minimum requirement for MHs is the movement detection, refreshment of registrations and address registering. If

independence of Address Management from mobility, i.e. handover, is striven for address registering can be removed from the minimal requirements.

If a particular design choice is to place minimal requirements on MHs, realisation of it becomes the manner in which the movement detection procedure is solved and the messaging format for refreshing network-layer registrations (Packet Forwarding state and additionally macro mobility bindings). If the consideration for movement detection is limited to network-layer procedures (as link layer solution are out of scope of this model and a implementation-specific) messaging format can be matched to the macro-mobility solution (i.e. Mobile IP) and triggering of Path Updates would be performed in BSs as already specified for Path Updates Solutions. Alternatively MHs would need to conform with the particular format of Path Updates messages.

Requirements for MHs PDI directly overlaps with the complexity and mechanisms for Handover Management. If the above-explained design strategy of minimal requirements for MHs is followed, design of Handover Management Solution becomes limited, as MHs are not heavily involved in the execution procedures. In such scenarios Handover Management would be limited to network-controlled, unplanned handover support with difficulties in interpreting old BS and no assistance from MHs.

Requirements for Global Internet Interfaces

A straightforward requirement for Global Internet Interfaces could be the arbitrary location and number of network gateways. As already mentioned in section 5.2.2.6 these can be administrative gateways without the mobility-specific functionality or one of them needs to serve as the mobility gateway (see Figure 5.6 and Figure 5.7 for host routes and cascaded tunnelling Packet Forwarding techniques respectively). As packets entering the network are usually prefix-routed to the mobility gateway (as the above Packet Forwarding paragraphs explain), in case of multiple administrative

gateways, packets would have to be routed to the serving mobility gateway before they can be forwarded downstream to the MH according to the packet forwarding setup. One solution to this arrangement is to have Mobility Gateway decoupled from the Network Gateway. However, in such scenario there might be a risk of non-optimal routing in the network if the route from the Mobility Gateway to MH's current BS traverses routers upstream towards to Network Gateway before being forward to the BS (particularly relevant for hierarchical topologies if there is no direct downstream route to current BS).

One particular packet forwarding setup which does not require coupling of administrative gateway functionality with the mobility gateway functionality is the *partial default prefix-based routing* technique where packets are prefix-routed inside the micro-mobility domain to the **address origin** hence not placing any gateway-requirements on the network setup (see Figure 5.8). A similar property can be observed for *"hard-state" prefix-based routing* technique (see Figure 5.9). This points out the direct relation between the Requirements for Global Internet Interfaces and Packet Forwarding Solutions.

Address Management

From a design perspective Address Managements, as the solution driver, can be one of the most influential PDI Solutions as far as the setup of the remaining modules (i.e. PDI Solutions) of the mobility protocol. A designer can consider the following general considerations when devising an address allocation procedure (here a particular approach is taken where a designer considers the deployment implications of different addressing methods):

- **Extent of the available address space for allocation to MHs**: Address space can be limited and controlled by the network. This may affect the decision for allowing

stateless[4] or statefull (network managed via an address allocation procedure, e.g. DHCP) allocation of addresses to visiting MHs. In typical stateless address allocating procedures, addresses can be configured by MHs either solely by MHs (IPv6) or assisted by a link-local Mobility Agent (IPv4). In statefull allocation of addresses these can be allocated by the network's allocation function thus effectively controlling the extent and type of allocated addresses.

- **Topological characteristics of the available addresses**: Continuing the previous point address allocation can be influenced by the topological characteristics of addresses which are to be assigned to visiting MHs. Should the address space be explicitly managed by the network, addresses can be deliberately void of internal topological relevance and only have the prefix which is externally routable to the micro-mobility network (HAWAII, Cellular IP, MMP). Such a setup for Address Management directly promotes the use of *host routes* techniques for the Packet Forwarding solution (see Figure 5.6). An opposite situations is the *cascaded tunnelling* techniques as the Packet Forwarding choice which would require topological correctness of addresses obtained by MHs but would also imply address changes during handovers (see Figure 5.7). Another example of assigning topologically correct addresses is the *partial default prefix-based routing* Packet Forwarding technique where allocated addresses are topologically correct but the address change is not required during intra-domain mobility i.e. handover (see Figure 5.8).

- **Dynamics of the address allocation**: Address Management design can be influenced by the particular question: should MHs be encouraged to retain the same address during the intra-domain mobility and handovers? Reasons for encouragement of address preservation during mobility could be easier determination of MH's identity and location for achieving interoperation with other protocols which use the

[4] Stateless and statefull terms are intended to define different methods of address acquisiton. These are similar in meaning to the terms used for IPv6 addressing procedures but are applied in a more broad

MH's address (e.g. QoS support, security). One the other hand, an operator may consider other reasons for allowing address changes as a consequence of intra-domain mobility (e.g. HMIP) by for example using other PDIs as design drivers.

The available choices of Address Management PDI Solution can be crudely summarised by the following design questions: which entity in the network controls address allocation, what are the topological characteristics of allocated addresses, does intra-domain mobility affect address change and what are the implicating of Address Management on other PDI Solutions?

As evident from the above, the most directly related PDI Solution to Address Management is Packet Forwarding, which as explained, directly influences Requirements for Global Internet Interfaces and so on. The reverse interdependence is also correct.

Since this interdependence between the mentioned PDI Solutions is inevitable, in order to make a tentative conclusion on a particular strategy for address management the initial design principle of minimal interdependence of PDI Solutions should be considered. This principle is the main reason for concluding that the features of Handover Management, which should not play any part in address acquisition/allocation procedures. Thus, mobility (i.e. handovers) should not automatically incur address change and this narrows down the choice of Address Management Solutions to allocation of topologically correct addresses without mobility-induced changes. And as mentioned, this maps into *partial default prefix-based routing* and could be adapted to *host routes* technique. For further specific elimination, deployment-subjective criteria can be considered.

Routing Topology

A mobility design can constructively process particular parameters that refer to the operational aspects of mobility protocols inside networks. This, as explained in

---

scope in the text.

section 5.2.2.8, relates the all aspects of the topological and static view of the network. Should Routing Topology be used as the design driver then the following issues constitute the framework for considering its impact on the rest of the PDI Solutions:

- **Recovery from link and router failures**: All mobility solutions tackle these problems. A particular network scenario may involve sparse population of routers with dense and mesh-type interconnection where failures of links are more tolerable than the failures of routers. In such scenarios *cascaded tunnelling* techniques offers an inferior solution because of the dependency on the setup and function of Mobility Agents (see Figure 5.7). Generally, *cascaded tunnelling* techniques place large dependency on routers i.e. Mobility Agents, whereas other Packet Forwarding techniques dynamically adapt to network changes. This can be taken as applicable to *host routes* packet forwarding technique since a failure of one link or router would typically spark re-creation of the tree via possible route/router. While this observation is mostly valid for mesh topologies in case of hierarchical topologies there are obvious risks of failures applicable to all mobility scenarios (due to existence of a single route to a particular entity in the network).
- **Routing state maintenance**: "Soft state" methods used by *host routes* techniques offers reliable testing of the real-time operational stability of the mobility protocol but incur overhead compared to other Packet Forwarding techniques which can rely on default mechanisms inside the networks, i.e. intra-AS routing protocols (e.g. OSPF, RIP). The protocol overhead due to route maintenance ("soft state", i.e. "keepalive" mechanisms) is examined in Chapter 4.
- **Network topology dependency**: Essentially, all mobility protocols are capable of running in any network topologies. While the *cascaded tunnelling* technique overrides the underlying network topology, *host routes* technique builds the forwarding tree according to the underlying network topology. This may be an important factor when determining the method for transmission of Path Updates, which create the tree.

Usually this point relates to targeting of Path Updates, whether they are addressed to the mobility gateway (MMP) or old BS (HAWAII) where in the latter case this may cause some not optimal tree establishment and transient looping during handovers as discussed in section 5.2.2.3.

Security

Security features can often have limited impact on development of the rest of the mobility functionality. This statement is valid assuming that securing of protocol's operations is usually performed once protocol mechanisms have been developed. Once such example is shown in section 3.4.3.6 where security mechanisms that could be deployed in MMP are described. However, some security features could have an impact on the rest of the mobility functionality. Taking into consideration the effect of a Security setup on the rest of the mobility features decision like extent of encryption of control messages can impact the Requirements for MHs and Handover Management especially if encryption of control messages is required over the wireless link. Another particular feature, which affects Path Updates and Handover Management, is separation of the two and triggering of Path Updates by the new BS after handover. Such as scenario can allow generation of network managed (and thus trusted) Path Update messages. Also, Security can require transfer of encryption keys from new to old BSs during handovers in which case Handover Management needs to include such a procedure.

5.3.1 Conclusion

The presented study of implications of using each PDIs as design drivers and considerations of their individual functionality in relation to the rest of the mobility setup follows the principle of modularity and minimal interdependence of PDI Solutions and proposes design choices for each PDI Solution. The first conclusion from the study of PDIs concepts shown in section 5.2.2 and their interrelations shown

in the previous section is that the identified PDIs impact different aspects of mobility design.

Path Updates, Handover Managements and Paging can create a backbone of the protocol in terms of flows and management of control messages, i.e. signalling. Hence, from a perspective of IP functionality needed for actual realisation of a mobility protocol these three PDIs can be considered as the most apparent features (provided they are chosen as needed functionality considering that Handover Management and Paging are optional features). This assumes that the elementary descriptive functions of any IP mobility protocol are typically the flow and management of control messages.

Packet Forwarding can also incur specific mechanisms for protocol operation. These do not relate to message transfers as the previous three PDIs but to specific installations of routing states in the network (e.g. Mobility Agents/FAs in the *cascaded tunnelling* case, or specific entries in routers in case of *host routes*).

PDIs such as Address Management, Routing Topology and Requirements for Global Network Interface mostly provide design constraints on remaining PDIs and their Solutions rather than inducing specific features in the signalling flow of the protocol. An example with Address Management PDI can be a requirement for continuous lifetime-long *care-of-address* for MH in a foreign network, which immediately makes the *Proxy-Agent Architectures* (i.e. *cascaded tunnelling*) and their relevant PDI Solutions unacceptable because of the inevitable change of *care-of-addresses* associated with the changes of BSs/Mobility Agents in the same network. PDI such as Security can usually be adapted to the specific mobility features.

Observing the separations of mechanisms of PDI and their Solution, the most straightforward PDI split is differentiation of Handover Management from Path Updates (since as already stated these are manifested in functional properties of a mobility protocol). This statement goes along some other research efforts in the Internet, which attempt to differentiate between **vertical** and **horizontal** signalling in

the network [49]. Vertical signalling corresponds to the distribution of control messages for facilitating refreshments of forwarding entries in mobility entities such as routers or alternatively Mobility Agents. These entities maintain routing entries for MHs inside the network. The vertical signalling concept, to a large extent, maps to the definition of Path Updates of the proposed PDI split. Horizontal signalling refers to all protocol mechanisms for facilitating changes of points-of-attachments (BSs or Access Routers depending on the terminology, assuming they are IP capable) and roughly corresponds to the Handover Management PDI.

Another conclusion from the analysis of interdependencies of PDIs, and their potential selections as individual design drivers, is that different mobility solutions can be constructed based on specific choices of primary PDIs and their chosen basic features. If minimal complexity in MHs were a particular design principle, this would emphasise minimal Requirements for MHs PDI and the resulting lack of dedicated features for Handover Management PDIs. One example of this setup is MMP where MHs are only required to run Mobile IP mechanisms over the wireless medium to BSs and where all signalling related to changes of point-of-attachment is performed by Path Update PDI Solution (MMP PDI split is analysed in section 5.5). Another example can be the solution for Address Management PDI for visiting MHs in a network. If addresses are managed and specifically owned by the network but not routable inside the network (i.e. only routable to the network as they would have a prefix belonging to a network's address prefix) then this would naturally promote *host routes* Packet Forwarding PDI Solution as the method for delivery of packets to MHs. In overall, the model presented in this section does not converge the mobility solution space into a single mobility protocol or a set of features for some PDIs. It offers a general model for development of mobility solutions by following "design threads" influenced by the selection of primary PDIs. A reflection of the principles of generic design model is detailed in the following section with the actual subjective specification of all design parameters, developed in the BRAIN project. As evident

from the following text and as mentioned in section 5.1 creation of the design model was aimed at producing a practical compromise between the available features of mobility protocols for a specific deployment scenario not a single "ideal" mobility solution.

5.4 Design of BRAIN Mobility Solutions

5.4.1 BRAIN Design Principles

The BRAIN project (and its follow-up project MIND[5]) [47][89] was managed by European Commission involving numerous telecommunication companies dealing with development of IP access networks (ANs) and their architectures. Design of IP mobility protocols(s) for deployment in the BRAIN IP access networks was one of the main design goals of the project. More on the project details, technical topics and author's involvement in the projects is given in section 1.6 and Appendix 2.

This section contains description of some of the analytical processes applied in the BRAIN project that lead to creation of BRAIN mobility solutions. As already mentioned initial work in the project was driven by the concept and application of the Evaluation Framework for initially evaluation the existing mobility protocols and organisation of research activity. For this purpose many of the mobility protocols were evaluated using the evaluation criteria, which is exemplified in section 5.2.2. Regarding the development of the final solutions for the BRAIN project a solution was striven for using the concepts of PDIs and their split described in the previous section[6][47][5][8]. As the previous section explains much of the design process is

---

[5] The results from the project are mostly related to research performed in the BRAIN project and its specific results further refined in the MIND project (follow-up of BRAIN). The MIND project had a larger scope of research such as multi-homing, ad-hoc and self-organising and moving networks and this is not related to the presented results. Further explanation on the scope of research is given in Chapter 1 and Appendix 2.

[6] Taken from the final public BRAIN deliverable 2.2 (section 3.1 in [47]): "In the BRAIN project, we have performed a critical analysis of them through an Evaluation Framework [3.1], extracted the key functionalities that must be realised [3.2], and have also developed a "BRAIN Candidate Mobility

concerned with selection of the design principles imposed by the network designer, which affect the mobility solution by influencing selection of primary PDIs.

The following explains the specific and emphasised design principles used in the BRAIN project [8](more on the entire set of design principles in the BRAIN project is given in [47]):

1. **Obeying the basic IP principles** assumed in this document and applicable to default IP networks. Some of these basic principles are maintenance of end-to-end IP connectivity between IP hosts (i.e. BRAIN IP networks are not intended to impede or alter the current communication models of the Internet), layered design (i.e. IP is the network layer with OSI relations (see section 1.2) with other functionality and layers)…
2. **A well-defined problem scope**: The problem scope was an IP access network where MHs move around from one BS (or BRAIN Access Routers in the particular BRAIN terminology) to another. From a pure IP perspective this AN, which is controlled by a single operator is a single administrative domain (intra-AS system and a micro-mobility domain) and it is of arbitrary size.
3. **Micro and macro mobility should be solved by separate protocols**: This follows as another corollary of the transparency principle. If macro mobility (i.e. MH moving between ANs) was to impact micro mobility, then the latter would be unable to work with alternative macro mobility protocols - which would violate the transparency principle. Further, since routing within the network is based on the locally assigned IP address, when the MH moves into another network it must obtain a new IP address. This conclusion is different from much of the existing IP mobility work, which tries and solves all mobility problems with a set of extensions closely coupled to some version of Mobile IP. Although this point was

---

protocol" that may be suitable in many scenarios." Evaluation Framework is presented in its complete from in section A3.3 of [47] and the deliverable includes analysis and solution for chosen PDIs such as Handover Management, Path Updates and Paging as explained in the following text.

considered important in the design process, it assumed Mobile IP as the current macro mobility solution because of its current use and status.

4. **Transparency principle**: This was the term used to capture the assertion that the basic goal of the AN is to make mobile wireless Internet access look like 'normal' access through wired infrastructure. For example, this means that the AN completely hides, from other fixed networks and correspondent hosts (CHs), the mobility and wireless aspects; these are only visible as performance impacts such as transient QoS variations [3] - just as occurs with other access systems. Additionally, the AN is expected to perform appropriate uplink and downlink forwarding of packets without changing or discriminating their contents, provided MHs are authorised to use the network.
5. **Impact of transparency principle on addressing**: The AN must allow a MH to get an IP address to use in communicating with CHs (again no assumption on macro-mobility mechanisms because the AN may be the 'home network' of MH), and to keep this address whilst it moves (because if mobility caused its address to change that would be visible outside the AN, violating the 'transparency principle'). Therefore the network must:
    a. assign addresses to MHs, which are globally reachable
    b. route packets to the MHs based purely on these locally assigned IP addresses.
6. **Modularity**: Starting from the initially identified Protocol Design Issues it was considered beneficial to re-group the functions they represent into (semi) independent modules of the complete mobility solution. Extent and benefits of the modularity are identical to the ones explained at the beginning of section 5.3. This especially relates to separation of flows of control messages already mentioned in the previous section refereeing to mechanisms of Path Updates, Handover Management and Paging [8][47].

7. **Scaling and Resilience**: These specific requirements on the micro-mobility solution are identified as recommended characteristics of final solutions rather than strict rules for determining primary PDIs. Resilience initially relates to Requirements for Global Internet Interfaces, where multiple network gateways should be supported, and Routing Topology, where there should be fast recovery in case of link or router failures. Scaling initially relates to Routing Topology: route aggregation should be supported. These should accordingly impact the choices of primary and other PDIs.

The following section presents the implication of this set of design principles by presenting a high-level platform for desirable mobility solutions in BRAIN IP access networks. This is then followed by a protocol proposed inside the project intended to satisfy the here-presented design principles and conform to the functionality of the high-level platform.

5.4.2 BRAIN Mobility Solution

After considering the identified design principles, their impact and relation to the functionality of associated PDI Solutions, the conclusion is that the BRAIN platform should be represented by placing the exact requirements on the final functionality of Packet Forwarding, Address Management, Handover Management, Path Updates and Paging (i.e. Support for Idle MHs)[8]. These functional requirements and the resulting scope for solutions are explained below for each of the chosen PDI. Observing the previously explained design principles the following implications on the PDIs can be stated:

- General Implications: this is related to the first 3 design principles (obeying the basic IP principles, a well defined problem scope and micro and macro mobility) which provide a background description of the BRAIN network functionality and can be assumed to provide the same functional platform for development of IP

mobility solutions as encountered in relevant outside IP mobility research (see section 1.4 for scope of research).

- Address Management: This comes as a direct consequence of the transparency principle and its implication on addressing where MHs are generally expected to retain their globally reachable addresses during their connectivity inside the network (naturally, change of address should be possible but not as a direct consequence of handovers).
- Handover Management, Path Updates and Paging (Support for Idle MHs): As already noted in section 5.3.1, these three PDIs form the apparent functional features of mobility protocols (i.e. control message flow). Hence modularity as the design principle mostly affects these three PDIs where the Handover Management is the interfacing sub-protocol for MHs connecting to BSs, Path Update refers to how network deals with internal updates and signalling (decoupled from Handover Management regarding the message flow) and Paging, which can also follow the rules of decoupled protocol parts for MHs and network-internal processes.

The setup of other PDI Solutions should either be deducted from the functionality of "the platform PDIs" and their interrelations shown in section 5.3 or left open to the particular design preferences. In the following, the modules are discussed from a high-level perspective where the engineering of their particular features and the extent of their functional independence are left open (more is given in [8]):

1. Address Management: Address allocation is independent from the other 'modules'. Address change does not occur as an automatic consequence of mobility inside the network. However, it can be triggered by a MH or the network, according to their particular preferences. This approach to Address Management contrasts with many other IP mobility solutions; one of its advantages is that the network operator can manage addresses in the manner that suits them. One realisation of this specification is that the allocated address is prefix routable inside the network (apart from being globally routable) and its **address origin** should be any possible entity in

the network: BS, any internal router or any network gateway. Such Address Management setup fits with *partial default prefix-based routing* Packet Forwarding technique as discussed in section 5.3 (with arbitrary address origin and the technique for routing from address origin to the current BS) and alternatively *host routes* Packet Forwarding technique (with arbitrary mobility gateway location). More on the specific choice of Packet Forwarding is given in point 4 below.

2. Handover Management: Separated Handover Management gives flexibility and freedom for development of other PDI Solutions (see previous section). Depending on the suitability to a particular deployment scenario handovers can be mobile or network controlled/assisted, independent/dependent on particular wireless scenarios...As mentioned there should be no direct relation and interdependence between the addressing and any stage of the execution of the handover sub-protocol. This also assumes minimal dependence on the Path Updates and annuls the minimal Requirements for MHs PDI Solution. **A particular design of a Solution for Hanover Management is shown in Chapter 6.**

3. Path Updates: Assuming that Handover Management protocol takes care of the temporary redirection of packets during the handovers, the importance of Path Updates with respect to *handover latencies* (i.e. packet losses) is reduced. Path Updates should be triggered when the BS transfer facilitated by the Handover Management is certain. The initiator of the Path Updates can be BS or MH. However, since MH is inevitably involved in the stages of the handover protocol and since handover is decoupled from the rest of mobility, it is beneficial to "hide" the process of Path Updates from MHs and trigger them from BSs (as mentioned in section 5.3 this option may also be preferred for Security reasons).

4. Paging (Support for Idle MHs): Paging Solution should run on its own when required, with minimal interactions with other mobility modules (PDI Solutions). Any solution should comply with the description of Paging as the design driver given in section 5.3.

5. Packet Forwarding (Routing): Packet Forwarding as a separate module (PDI) is not separable from the rest of mobility as clearly as other above presented PDIs and is not identified as a direct consequence of the design principles. However, it involves an important design decision, which according to the presented platform can be considered independently. These decisions involve further issues than the requirement on Packet Forwarding derived from the Address Management PDI identified above and specified in section 5.3. Packet Forwarding considering the above setup mostly relates to the way in which packets are forwarded from the address origin or mobility gateway in case the respective choices for Packet Forwarding techniques influenced by the requirements on Address Managements PDI are *partial default prefix routing* (with any techniques from address origin) and *host routes*[7] (with arbitrary mobility gateway). Due to the options possible for support of Packet Forwarding PDI and scaling and resilience design principles shown in the previous section choices of address origin and number of Global Internet Interfaces (i.e. network gateways) induces various design choices. Regarding *partial default prefix-based routing* the choice of the techniques from the address origin to the current point-of-attachment becomes a design choice between installing a host route, tunnelling or creating a "hard state" prefix-based routing states. The BRAIN project was not able to reach a solid consensus on whether host routes or tunnelling is better, views depended on the subjective interpretation of design principles and their implementation implications (this is taken from section 4 in [8]). Hence, the above discussion gives explanation to some of these issues (more on this is given in [8] and section 3.4.2. in [47]).

The presented set of PDI Solutions and the options specified for their realisation in a mobility protocol presents the BRAIN mobility platform for designing desirable Mobility Management protocols for BRAIN networks. The research on mobility solutions was concluded with the platform from whereon it was left to a particular engineering solution to comply with the specified functionality. One such solution is

[7] Some restriction on address origin may exist

the mobility protocol BCMP explained in the following section, which is followed by a section showing how the mechanisms of BCMP comply with BRAIN design principle and the identified BRAIN mobility solution. As the word Candidate suggests, BCMP is a possible practical realisation of the BRAIN mobility solution. Any additional solution, which complies with the specific platform, could also be accepted as a desirable outcome. This point goes along the one mentioned in section 5.3 where the aim of the design model presented in this chapter is not to derive a single mobility solution but to reach a practical compromise between the available solutions and to conform to particular design principles.

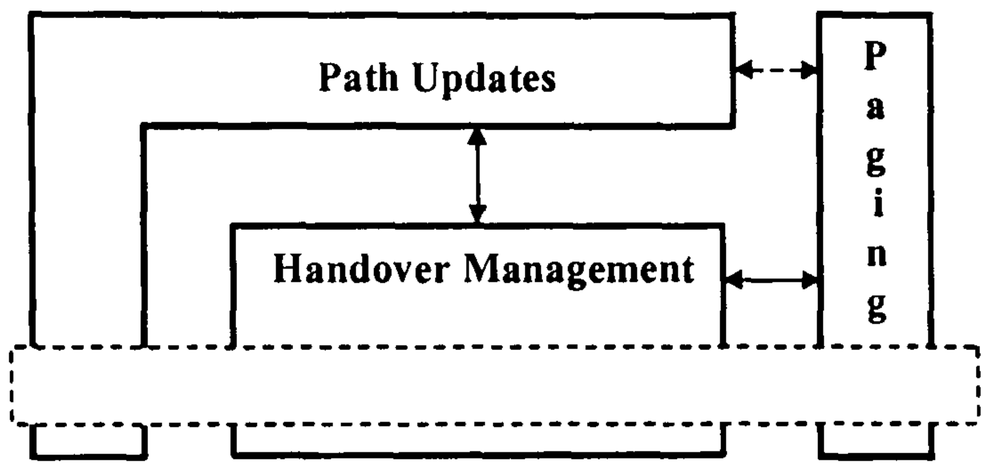


Figure 5.10. Main Functions of BRAIN IP Mobility Solution

The main functions of the BRAIN IP Mobility Solution are presented in Figure 5.10 taken from [47] and related to the separation of the three PDI, which form the flow of control messages of the mobility protocol. The figure does not include more "descriptive" PDIs: Address Management and Packet Forwarding (Routing) explained above and in [8][47] with their specific requirements for BRAIN networks.

5.4.3 BRAIN Candidate Mobility Protocol – BCMP[8]

This section gives a short overview of the main features of BCMP. As already noted in this chapter, BCMP is proposed inside the BRAIN project as a mobility solutions that satisfy the main functions (PDIs) of the BRAIN IP Mobility Solution presented in

the previous section. **BCMP is not attributed to the author** (this is further explained in section 1.6). BCMP is included for descriptive purposes and as one of the results of the work in BRAIN project. In addition to the following description of the essential features of BCMP, the next section further analyses the features of BCMP against the design model presented in this chapter and the BRIAN mobility platform derived from the PDI split. More on the specific features of BCMP is given in [47] [61] [8].

BCMP operation is targeted at IP micro-mobility domains where a macro mobility protocol is not considered but applicable as the next section discusses. The central entity of the Packet Forwarding setup is the Anchor Point (ANP), which is located in the network (BRAIN Access Network or any other micro mobility domain) where BCMP may be deployed. ANP is also the "address origin" meaning all addresses allocated to MHs in the network are globally and internally prefix-routable to the ANP. There are no restrictions on the number and location of ANPs in a network: an ANP can be any internal router, network gateway or BS. All downlink packets sent to MHs are routed to their serving ANP from where they are tunnelled to the serving BS.

MHs run a specific Handover Management sub-protocol to connect to the network, maintain connectivity and perform handovers. MHs are not required to execute any other internal procedures apart from the ones specified in the Handover Management[9] protocol and involving MHs and BSs. The particular mobile-controlled/network-assisted planned and unplanned Handover Management protocol adopted for BCMP complies with the Handover Management PDI Solution explained in the previous parts of this chapter and in accordance with the BRAIN mobility platform explained in the previous section (see section 3.2.2 in [47]). The whole design and features of the Handover Management PDI Solution which is adopted for BCMP and recommended for other micro mobility protocols is shown in Chapter 6 and is not covered in detail in this chapter.

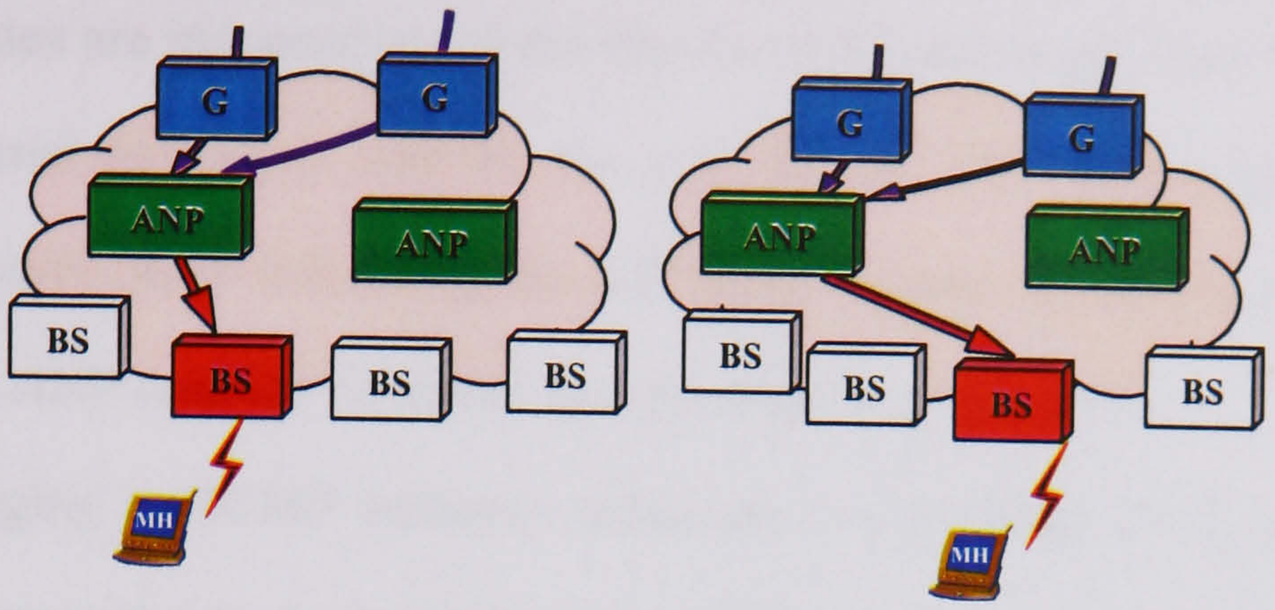


Figure 5.11. BCMP Packet Forwarding Setup for a MH handover

The main BCMP procedures are:

▪ **Address Management:** During the login MH is assigned an address belonging to the address space of ANP. The selection of the serving ANP can be performed automatically by the BS to which the MH attaches or by another controlling entity in the network which can exercise any pre-set policy for address (thus ANP) allocation. The network (either ANP or another controlling entity in the network) executes the security (e.g. AAA) procedure to identify and authenticate the MH and allocates a globally routable IP address from the address pool belonging, and prefix-routable, to ANP. The session key (for security purposes) and IP address are sent to the MH as a login response. Change of serving ANP is possible but not as a mobility-imposed feature since MHs retains the serving ANP during handovers. Address change may demand ANP change since addresses "belong" to ANPs (this constitutes a macro mobility event). Address change can be requested by MH or the network in response to a network management event.

▪ **Handover Management and Path Updates**: The protocol deploys the Handover Management procedure according to the design shown in Chapter 6 for both planned and unplanned handovers for communications of MHs and BSs. Internal signalling,

---

[9] The specific Handover Management sub-protocol (PDI Solution) adopted for BCMP and designed in Chapter 6 was also referred to as MH-BRAIN AR (BAR) protocol in the particular BCMP terminology [47][61].

i.e. Path Updates, are triggered during the handover process but not controlled by MH as Path Updates are independent of the Handover Management. Path Updates relate to network-internal messages: sent by the new BS to ANP diverting the traffic after handovers and/or ANP informing the old BS to release the temporary tunnel to the new BS. The ANP updates its tunnel end-point during handovers.

▪ **Paging**: Paging in BCMP achieves reduction in signalling. Packets destined for an idle MH are tunnelled to the last known BS, which initiates the paging process.

5.4.4 Analysis of BCMP

This analysis of BCMP is included to show how its features comply with the design model presented in this chapter used in the construction of the BRAIN platform for solving mobility (section 5.4.2) and considering the particular BRAIN design principles (section 5.4.1). Adherence to the "transparency" principle can be explained considering the following points:

- As a natural consequence of preservation of allocated addresses, mobility of MHs is hidden from the external networks hosts
- The Hanover Management protocol is the interfacing protocol for MHs connecting to the network. The rest of the mobility mechanisms inside the network is hidden from the MHs. In BCMP, the setup of message transfers (in particular the Handover Management) is such that the location of serving ANP or any additional information about it is kept hidden from MHs. In simple terms, this means that the MH acquires an IP address during the login phase and it is only required to run the Handover Management protocol without any further mobility-related intelligence thus keeping the inside of the network invisible to the attaching MHs.

BCMP executes AAA-based control at the initial login. In practical terms, upon the initial message exchange between BS and MH, the BS consults a local AAA-entity and the address allocation entity in the network (in case the address allocation is not

automatic) before the appropriate ANP is selected. Furthermore, BCMP deploys a reverse tunnel from the BS to the ANP for uplink transmissions by MHs. This feature enhances the "transparency" of the protocol and offers a useful functionality for other mechanisms of the network. In abstract terms, this causes all traffic flowing to and from the MH to appear as if it terminates and originates at the serving ANP.

Effectively, the template BRAIN platform for solving mobility and BCMP, as one example of its practical realisation, is a stand-alone micro mobility solution. It does not inhibit the use of any macro mobility protocols such as Mobile IP and can support other macro mobility mechanisms such as SIP. The stand alone property of BCMP is considered important for BRAIN connectivity scenarios where the connecting MHs may use the network as their home network.

MHs are allocated globally routable IP addresses. Considering BCMP, MHs are allowed to change the obtained address not as a mobility-imposed necessity but as the controlled option for achieving optimisation of packets flows or for provisioning some other architectural aspects (QoS, ingress filtering, multi-homing). Also, MHs are allowed to request more IP addresses, even subnets, provided they satisfy the necessary requirements for such actions (again they requirements often depend on other aspects of the network architecture).

Modularity and the concept of PDI split is obeyed in the BCMP design. BCMP overall setup can be decomposed to reveal the individual protocol elements of the adopted template, i.e. BRAIN platform for solving mobility. Separation of the air interface protocols and the network-internal protocols directly maps into the distinction between the Handover Management protocol and Path Updates (with Paging as independent part of the overall functionality). The addressing setup of the protocol does not impose requirement on the Handover Management and Path Updates processes. Location and number of ANPs in a network is not restricted. Due to the Handover Management protocol, location of ANP (conceptually representing the

"cross over" routers) is not expected to influence packets losses during handovers. This aspect is further analysed in Chapter 6.

The Packet Forwarding technique used in BCMP is *partial default prefix-based routing* up to the address origin, which is the ANP, from where tunnelling is used from ANP to the serving BS.

As pointed out in section 5.3, modularity allows for easier enhancement and evolution of the mobility solution through modification or complete change of some protocol parts (PDI Solutions). This particular property has been the driver for the protocol refinement work on BCMP performed in the MIND project dealing with its development.

BCMP interoperates with other architectural components such as QoS, Radio Resource Management, security and multi-homing. For more details refer to the overall IP architecture of the BRAIN network given in [47].

The initial consideration for resilience of mobility solutions is that the specific mobility-imposed resilience mechanisms should be minimised and mostly assisted by the embedded support in the network. The conclusion is that the resilience of BCMP comes from the property of the protocol, which does not rely on the specific installation of routing entries and thus overcomes any failures in the network using the default mechanisms for the internal routing setup in the network. ANP are the "weakest points" of the protocol, hence ANP changes are needed should the serving one fail. The utilisation of the network internal methods (because MH's address is prefix routable to the ANP from where the tunnel is again prefix routable to the serving BS) provides a natural support for multiple network gateways. BCMP does not inflict any routing requirements for the ingress or egress points of the network, hence allowing for arbitrary numbers of administrative network gateways.

This again allows for better scaling of the protocol as the process of assigning a particular ANP to MHs through the allocation of IP addresses can be further controlled by the network in a traffic-engineering manner. Additionally, utilisation of

the internal routing as the only routing mechanism of the protocol mitigates risks of bottlenecks (specifically imposed by the mobility support) and provides a sufficient level of confidence for scenarios of dense population of MHs. BCMP setup is applicable to any network size.

5.5 Application of the Generic Mobility Design Model for enhancing MMP

All already pointed out in the previous parts of this chapter the generic mobility design model can be applied not only for constructing novel Mobility Management protocols as shown in the creation of BCMP but for enhancing and adjusting the operations of existing mobility protocols such as MMP. Analysis of MMP's PDI Solutions in MMP is covered in section 5.2.2 and compared to other mobility protocols chosen. The analysis is void of any specific and subjective deployment considerations (as applied in the BRAIN project) and generally evaluates each MMP PDI Solution against the equivalent functionality in other chosen mobility protocols. As already noted in section 5.3 besides the construction and adaptation of mobility solutions for particular deployment scenarios (as applied in the BRAIN project), the work presented in this chapter can be used for assessment of a mobility protocol's performance compared to other relevant solutions (this is done for MMP in section 5.2.2), for increasing the modularity of the mobility protocols and thus achieving the benefits identified in 5.3 such as easier evolution of the design. Hence, MMP can be further analysed considering modularity as the design objective (default MMP was already analysed in the BRAIN project [5][47] with other relevant mobility protocols). This can give more detail on the properties of each feature of MMP in relation to the PDI split and their interrelations. A specific application of this study is exemplified in

possible integration of new Handover Management Solution as the one designed in the BRAIN project and presented in the next Chapter.

Related to the PDI split and MMP the following can be observed and stated for each particular PDI:

- **MMP Packet Forwarding PDI Solution**: MMP uses *host routes* techniques for installing routing "pointers" in the micro mobility domain. This is facilitated by CBT multicast protocol. The setup is the core functionality of MMP since it enables routing based on the multicast CoA. Packet Forwarding directly influences Address Management and Path Upadates since these are integral parts of CBT routing and demand a multicast addresses (i.e. care-of-addresses) and CBT control messages for installing host routes in the network. The protocol specification in Chapter 3 (section 3.4.3.2) also includes a Support for Idle MHs, which closely interacts with Packet Forwarding setup of CBT. Furthermore the specific Packet Forwarding techniques induces the solution for Routing Topology based on a CBT Core being the mobility gateway and CBT-specific "soft state" mechanism. In addition as indicated in section 5.3 *host routes* typically require coupling of network (administrative) gateways and mobility gateways (i.e. CBT Core in MMP). As far as Packet Forwarding is concerned, CBT could be replaced with a functionally identical but CBT-independent routing technique, which would release the dependency on specific CBT control messages and requirement for specific multicast CoA for identifying MHs. Such a solution would allow independent messaging and addressing solution in the networks, which would be left to designer/operator preferences.
- **MMP Path Updates PDI Solution**: As pointed out in the previous paragraphs specification of MMP explained in Chapter 3 assumes CBT messages for installing internal routing information in the network. In addition, there are Mobile IP messages transmitted by MHs and decoupled from CBT Path Updates, i.e. Mobile IP messages transmitted by MHs trigger CBT control messages in BSs. This shows that the default

Handover Management in MMP and Requirements for MHs do not clash with MMP Path Updates, which is a beneficial aspect of default MMP functionality and allows for easier evolution and enhancement of Handover Management as Chapter 6 explains. One particular message used in MMP is the *MMP Instruct* message which is a handover triggered messages which in turn triggers tearing down of the routing state i.e. "negative" Path Updates. The design of Handover Management and its potential integration with MMP would remove the need for *MMP Instruct* message (see Chapter 6) thus achieving interdependence of the PDI Solution. Following the conclusion from the previous paragraph dealing with Packet Forwarding if CBT is replaced with an arbitrary routing model the functionality and interrelation of Path Updates would remain unchanged but with arbitrary control message structure not dependent on CBT specifications.

- **MMP Handover Management PDI Solution**: The default Handover Management solution in MMP requires MHs to transmit extended Mobile IP messages where the only form of support from the network is the advance registration feature explained in 3.4.3.3. The previous paragraph on Path Updates explain how Handover Management can achieve independence from Path Updates and this is used as a basis for discussion on integration of the proposed Handover Management Solution shown in Chapter 6 which can also be adopted to MMP. The proposed Handover Management is the dominant factor in determining packet losses during handovers. However, the complexity of the proposed Handover Management protocol induces significant Requirements for MHs, which in the default MMP specification were minimal and related to Mobile IP functionalities.
- **MMP Support for Idle/Host/Paging PDI Solution**: The default MMP support for Idle Mobile Hosts interoperates heavily with Packet Forwarding and Path Updates and places requirement on mobility gateways i.e. CBT Core (Requirements for Global Internet Interfaces). As recommended in the previous discussions in this chapter

(especially sections 5.3 and 5.4.2) creation of a stand-along paging solution is desirable and in case of MMP this would release the interdependence of Paging with Packet Forwarding and Path Updates. Design of a Paging PDI Solution is out of scope of this investigation.

- **MMP Requirements for MHs PDI Solution**: The default features of MMP place minimal requirements on this particular PDI Solution. However, since the Handover Management protocol design in Chapter 6 is recommended for enhancing MMP this places significant requirements on the functionality of MHs since the dominant property of the Handover Management protocols is that it is mobile-controlled. If minimal Requirement for MH are striven for than the actual MMP design in section can be retained since MHs are expected to execute Mobile IP mechanism without any additional MMP-specific signalling.
- **MMP Requirements for Global Internet Interfaces PDI Solution**: The *host routes* Packet Forwarding technique used in MMP requires coupling between the network (administrative) gateway and the mobility gateway (i.e. CBT Core or simply Gateway if dependency on CBT is avoided as previous paragraphs propose). This could be partially fixed by placing the mobility gateway of MMP in arbitrary locations in the network and having the network gateway (or any of them if more than one is deployed) route packets inside the network to the mobility gateway from where they can be forwarded using MMP's *host routes*. This could be possible since downlink packets are addressed by CH or HA to Gateway's IP address before they are forwarded along the established routing tree.
- **MMP Address Management PDI Solution**: MMP deploys multicast CoAs as the default specification of the protocol uses CBT multicast routing. One benefit of the overall MMP setup is that the address change is not needed during handovers and it is not a mobility-imposed mechanism. As the previous paragraphs recommend, if dependency on CBT is avoided the current requirement for a multicast CoA could be

replaced by any unicast or multicast IP address since the *host routes* technique of Packet Forwarding can use topologically irrelevant addresses inside the micro mobility domain.

- **MMP Routing Topology PDI Solution**: MMP reliance on a single mobility gateway can be decoupled from the reliance on a single network (administrative) gateway by following the recommendations of the paragraph dealing with MMP Requirement for Global Internet Interfaces. In addition should dependency on CBT be avoided the default "soft state" mechanism should still be used to provide recovery from link and router failures. Network dependency is overcome in default MMP by adaptable route creation. Path Updates are addressed to the gateway during handover not to the previous BS hence assuming shortest path routing to the Gateway the formed tree is always representing the shorts routing distance between the Gateways and BS.

- **MMP Security PDI Solution**: MMP default security features are limited to the acceptance of any underlying security support and reusing the authentication procedures offered by Mobile IP. One beneficial feature of MMP Path Updates is that they are triggered during handovers by BSs (not MHs) so the critical part is the MH-to-BS messages of the Handover Management protocol. While the initial login can be solved by AAA-alike procedures, transfer of session keys to new BS can be facilitated via the context transfer procedure of the Handover Management protocol as Chapter 6 explains.

As already pointed out the Handover Management protocol presented in the next chapter is integrated with BCMP but is free of any dependency on specific Path Updates and can be integrated with MMP. Model for integration of Handover Management and enhanced MMP is given in the next chapter with performance analysis which are relevant to any protocol which deploys the proposed Handover Management PDI Solution.

If MMP is to be adapted to the BRAIN mobility solution, this cannot be done in its form presented in chapter 3. This was generally concluded for all mobility protocols in the BRAIN project resulting in the mobility solutions presented in the previous section (section 3.6.1 in [47]). MMP would have to be enhanced with a Handover Management protocol as the one proposed in the next chapter and a new solution would be required for Support for Idle Hosts. This would achieve the separations for Path Updates, Handover Management and Support for Idles MHs PDIs as required in the BRAIN mobility solution. In addition, the exact requirement for globally routable IP addresses for MHs would mean that the multicast care-of-addresses would need to be replaced with the addresses allocated from the "address origin" in the network which can be any router including BSs, internal routers or network Gateway. However, the techniques of CBT multicast routing could still be applied as a "host route" packet forwarding technique from the "address origin" where Path Updates are separable from Handover Management (as already the case in MMP).

CHAPTER SIX

Design of Handover Management Protocol

Chapter Overview

This chapter continues the study shown in Chapter 5 where mobility functionalities are divided into PDI Solutions. The focus of this chapter is in describing design of Handover Management PDI Solution designed in the BRAIN project and intended as a stand-alone sub-protocol of adequate mobility solutions. The Handover Management protocol is designed as a generic solution for micro mobility protocols but is tested in BRAIN/MIND projects as an integral part of BCMP. Model is also given for its integration with MMP, which can be adjusted to the PDI split as explained in the previous chapter. The study indicates particular requirements for Handover Management solutions, analyses existing Handover Management PDI Solutions in relevant mobility protocols and creates a specific mobile-controlled/network assisted unplanned and planned Handover Management protocol framework based on the devised development strategy.

6.1 Analysis of the Handover Management Protocol Design Issue

As indicated in the previously chapter, one of the main conclusions of the generic mobility design model is that some Protocol Design Issues can be separated to such an extent that their associated Solutions actually stand out as separate sub-protocols and can hence be designed separately. This statement almost completely applies to the split between Path Updates and Handover Management mainly because of the distinctive separation of the two PDI Solutions, which are both based on exchanges of control messages as their elementary features. When such one-dimensional similarity exists between these two PDI Solutions it is easier to define the exact boundaries between the mechanisms of each. As explained in section 5.3, following the primary objective of the Generic Mobility Design Model, which is maximum possible functional independence of PDI Solutions, there should be no direct relation of Handover Management and Address Management. This particular requirement is accepted in the design of BCMP and the adoption of the PDI split for MMP and presents an novel approach in designing handover facilitating sub-protocols in the context of micro mobility.

Going back to the functional similarity between Handover Management and Path Updates an illustrating example of an opposite situation may include analysis of the interdependence of Path Updates and Address Management. While the two PDIs may either be dependent on or independent of each other, essentially, their straightforward separation cannot be considered from a practical design perspective because their Solutions are realised in functionally different PDI Solutions: Path Updates refers to the message "mechanics" while Address Management is an issue which affects the functionality of many other PDIs but typically does not itself contain an exclusive set of protocol mechanisms to solve it.

Taking another case of Path Updates "versus" Handover Management, a mobility protocol cannot be complete without either of the two, even if the protocol features corresponding to the Handover Management PDI Solution have rather basic features, as it is the case with IP mobility protocols mentioned in section 5.2.2.3. The separation of Path Updates and Handover Management enables greater flexibility of development and deployment of mobility protocols because it enables creation of generic and scalable handover management procedures along with other general benefits of modularity mentioned in section 5.3. At the same time, this approach eases requirements for a generic Path Updates Solution. Recollecting the default MMP design in Chapters 3 and 4, the potential benefits of MMP's deployment may depend on the setup of particular networks. When facing such trade-offs, having the choice of a greater flexibility in deploying optimum Path Update Solutions would greatly improve the generic nature of IP mobility protocols. Basically, the "interfacing" PDI such as Handover Management (and the Requirement for Mobile Hosts) would actually remain common for every particular network, while the other PDI Solutions such as Path Updates would be deployed based on the operator's preference in a particular network. Such a setup would allow the Handover Management protocol to be "glued" with the appropriate Path Updates protocol and other PDIs, with no additional requirements for MHs, thus making the deployment of those protocols more flexible. At the first glance, the separation of Handover Management and Path Updates should be a straightforward process, i.e. processing of Path Updates should be triggered after the completion of the handover. In practice, Handover Management should ideally involve the three entities, which are involved in the transfer of MH's connectivity from one point-of-attachment to another. These entities are: the MH, and the involved BSs, i.e. the new and the old BS. Thus from perspective of Path Updates (which as indicated should start after the completion of the handover), a handover is completed when the transfer between the old and to the new BS is finished (or in more practical terms, when the handover to the new BS is imminent).

This chapter contains design steps for development of Handover Management PDI Solution intended as a compliment to any mobility protocol, which conforms to the PDI split. This Handover Management PDI Solution is developed in the BRAIN project and independently presented as a recommended solution for this aspect of BRAIN mobility functionality. This is also analysed in section 5.4.2 where BRAIN mobility solution is presented and the solution given based on the identified primary PDIs and requirements for their functionality[1]. Based on this, the remainder of this chapter includes analytical processes that constitute the main design decisions applied in the development of Handover Management PDI and the resulting framework for realisation of the features. This framework presents the "*functional requirements for handover signalling*" focusing on the expected features of the final Handover Management PDI Solution, which can be relative to the implementation scenario, e.g. signalling formats and flows and underlying wireless link layer solution. One such realisation is exemplified in the specification for the Handover Management PDI Solution shown in Appendix 1. Finally, Handover Management properties are examined via its performance analysis conducted in BRAIN/MIND projects as an integral part of BCMP. In addition, integration of Handover Management PDI with MMP is discussed.

6.1.1 Handover Management Design Decisions

The idea of allowing additional handover support for MHs has been addressed in HAWAII and MER-TORA although as an integral part of the mobility protocol not as independent module for solving handover issues. As explained in section 5.2.2.3 the

[1] From the BRAIN project results [47] (initial paragraph of section 3.2.3 Conclusions on Handover Management]): *"In this section, we define the scope of our handover procedure and present a layout of the proposed handover scheme. The design is based on an analysis of functional requirements for the handover signalling, which have been refined after analysis of existing IETF handover protocol proposals (for more information, see annex A3.1). A basic requirement has been that the protocol should be adaptable to various Path Update schemes (which could be derived from existing micro-mobility protocols that have been adapted to conform to BRAIN design principles). A particular adaptation is presented in section 3.6."* (Author's note: Section 3.6 contains description of BCMP).

proposed solutions do not present a complete evolved approach, which could ultimately be integrated in a fully functional Handover Management protocol. Thus, a complete Handover Management protocol should be designed by capitalising on the experiences of these current solutions.

Again, the starting point of the handover protocol design is the reference to the handover-related protocol procedures of Mobile IP protocols both for IPv4 and IPv6. In both versions of Mobile IP, the new access points (FA in Mobile IPv4 in case of FA-CoA, IP access router in IPv6, referred to as BS in the following text) are discovered through a form of **network layer movement detection** mechanism (see sections 2.3.1 and 3.6.2). Generally, this is achieved by receiving network layer indications (in the form of advertisement messages) from new points-of-attachment (i.e. BSs). An additional step can be included to inform the old BS about the new BS's address in order to instruct it to forward packets to the new address until the HA or CH have been updated with the new location of the MH (by instructing the new BS to send an "update" message to the old BS). This temporary packet forwarding is achieved through an IP tunnel created between the old and the new BS. In Mobile IPv4, this is accomplished via the Route Optimisation (see section 2.3.1) whereas in Mobile IPv6 similar extensions are embedded in the base protocol using binding messages which can cause the old BS to act temporarily as a HA. This handover can be classified as an **unplanned handover** since the handover assisting procedures, represented in the protocol steps needed to set up the forwarding tunnel from the old to the new BS, are performed after the handover execution, that is, upon the attachment of the MH to the new BS. Mobile IP extensions for handover support achieve the primary goal of reducing the *registration delay*, which is primarily a product of the Path Update procedures showing the tight coupling and interdependence of the two PDIs as regards the handover procedures and performance. However, they fall short of achieving seamless handover because:

a) There is an inevitable delay associated with any possible movement detection procedures, which rely solely on network layer mechanisms (i.e. advertisements...). Depending on the underlying link layer technology, the interval between the establishment of the network layer connection between the MH and the new BS and the release of the previous connection at the old BS can be large enough to incur packet losses. This property arises from the (sometimes) inevitable break-up of the connection to the old BS (either explicitly by disconnection, of implicitly when the connection to the old BS is lost and a new attachment-discovery procedure is initiated) and the consecutive attempt to receive a new **network layer** advertisement message from the new BS. Hence, all network layer movement detection procedures are in reality not utilising the link layer information about the reachability of the new point-of-attachment. Actually, they await (or solicit) the reception of a network layer advertisement although the link layer connection with the new point-of-attachment is active[2].

b) The reactive nature of the Mobile IP Handover Management procedure incurs delays during the establishment of the forwarding tunnel between the new and the old BS. This is because MHs set up the tunnel from the old BS after they have attached to the new BS. The time required for the updating message to reach the old BS directly corresponds to the handover delay, which can often be sufficient to cause packet losses.

Generalised IP Handoff framework [49] is another research effort proposing a separation of the handover protocol from the rest of the mobility functions. As indicated in section 5.3, this approach coincides partly with the differentiation of a Handover Management PDI in the Evaluation Framework. Generalised IP Handoff provides a template for the flow of messages between MHs and access entities (BSs)

[2] The arguments of this paragraph are laid out to indicate the shortcomings of the Mobile IP movement detection algorithms. The Mobile IP procedures are heavily generalised and, in fact, different practical adaptation of those Mobile IP schemes can be deployed. The intention was to indicate the inefficiencies of network-layer Handover Management procedures and thus Mobile IP examples are used as reference solutions.

in networks. Various schemes have been proposed [50][51][52][53][54][55][56][57][58][59][60] deviating from the initial concept of Generalised IP handoff by proposing new messages and suggesting different platforms for their integration with the rest of mobility. However, the principal goal is still common, i.e. reduction (or rather a complete elimination) of packet losses during handovers[3], and their concepts of protocol interactions are largely similar (see Figure 6.1).

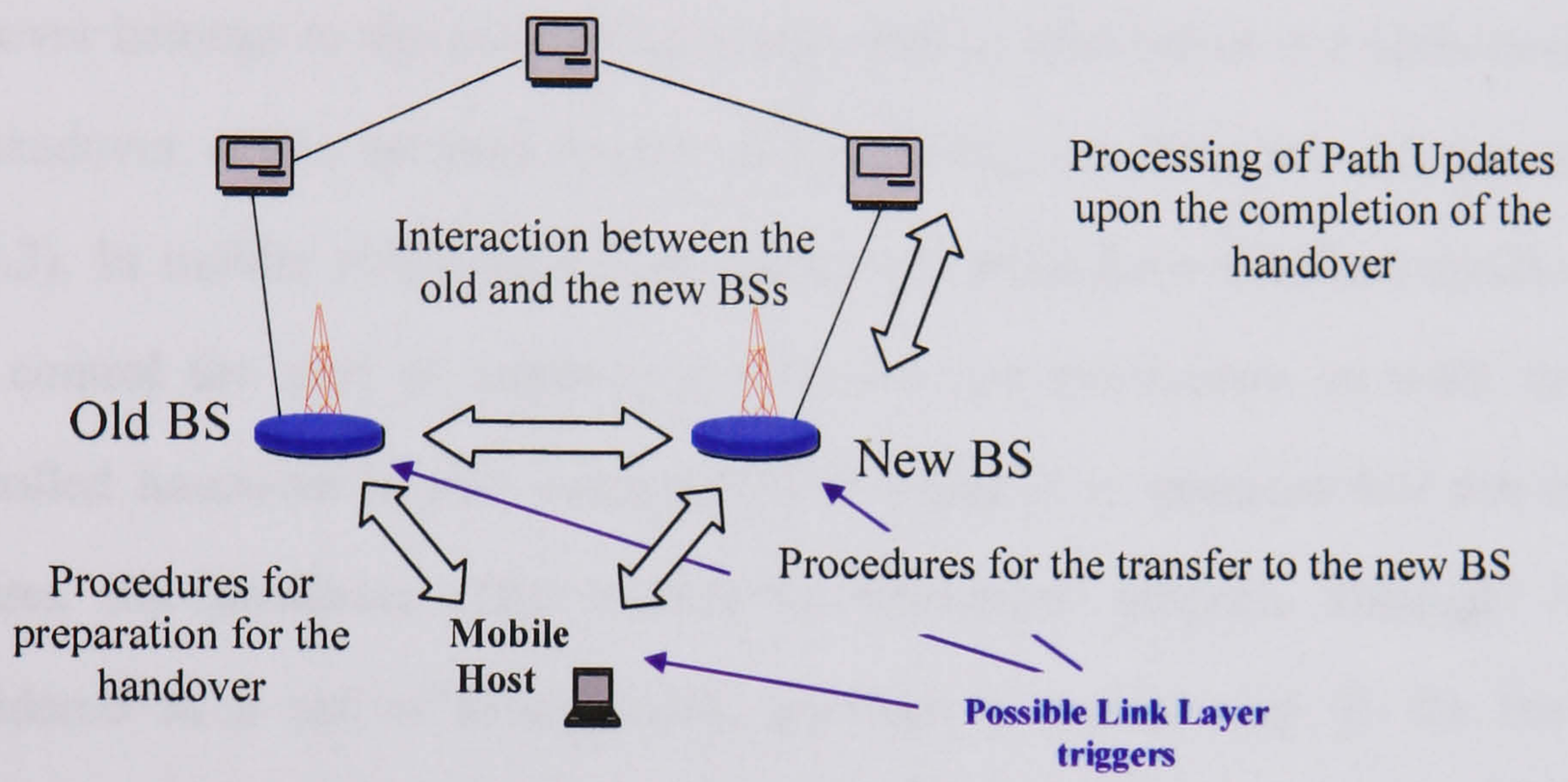


Figure 6.1. General outlook of possible steps in Handover Management. The inclusion of the Path Updates is only relative to the PDIs split proposed here.

Handover protocols can be analysed based on the following design decisions[4], which are, in different forms, present in all proposed solutions:

1. **Support for planned and unplanned handover:** As indicated in section 5.2.2, in planned handovers, the necessary signalling is assumed to take place (initiate) before the handover. A typical practical interpretation of this definition is that in a planned handover, signalling would take place before the MH loses its connection with the old BS, that is, as soon as it determines that a handover to a new point-of-

[3] This is an ongoing effort mostly conducted by IETF design teams. Recently, the proposals for IPv4 and IPv6 have converged into a single proposal per each version of IP, [59] for IPv6 and [60] for IPv4. However, various research conclusions are present in all schemes.

[4] Instead of individually explaining all the handover protocols referenced a general overview of the collective research issues is presented. This is due to the vast number of the schemes available and often the provisional and dynamic nature of the IETF proposals.

attachment is imminent. Most of the schemes proposed concentrate on providing models for planned handovers. This is a logical progression of the initial unplanned handover approach, proposed in Mobile IP, because it eliminates the delays mostly related to the lack of "planning" of the movements of MHs to new BSs. However, the general trend is still to consider the unplanned handover case but mostly as an alternative to the planned handover.

2. **Handover decision-making processes:** The final decision on the execution of the handover belongs to the controlling entity, that is, whether or not MHs decide on the handover, or the decision is up to the network, i.e. old or new BS (see section 5.2.2.3). In mobile (MH) controlled handovers, MHs have the final decision and thus control the start of handovers. Although not extensively covered, network controlled handover is also proposed [54], where it is assumed that the old BS initiates the handover. The underlying link-layer support, although largely considered as a set of assumptions, provides a crucial step in the handover decision-making in almost all proposals by providing for an initial "spark" for the handover, referred to as the **link layer trigger**. This can be related to the network layer movement detection procedures previously mentioned, which do not utilise possible knowledge of the proximity of a new point-of-attachment based on the link layer indication but require network layer advertisements. As mention before in this section, this is one of the primary causes for Handover Management-related delays. The features related to different link layer technologies can be summarised in the following way:

    a. Make-before-break and break-before-make, referring to the possibility of achieving concurrent connectivity to the new and the old BSs. The former provides for such a scenario, the latter does not.
    b. **Link layer triggers:** This refers to the particular link layer event, which causes an invocation of the network layer handover protocol steps. Cases are: source link layer triggers (at the old BS), target link

layer triggers (at the new BS) and link layer triggers at the MH. The specific choice of the link layer trigger depends on the handover strategy. A source trigger at the old BS can also assume that MHs initiate the process in a preliminary link layer-based exchange so that the old BS can use the link layer indication from the MH and then send the first handover message (network layer message). In the case of the target trigger, the new BS sends the first handover message to MH (target triggers can again be provoked by a MH via a link layer exchange). If the trigger is at the MH then the MH is required to transmit the first network layer handover message. Therefore, link layer triggers are not decision-making processes, but only events, which invoke the first handover message. The lack of network layer signalling for handover initiation (because of the link-layer triggers) leaves the discovery of the new BS to unspecific mechanisms (naturally related to underlying link layer). The only exception is the make-before-break solution adopted in [50], which does not rely on link-layer triggers, since simultaneous connectivity is achieved with the old and the new BS (make-before-break case). There is no direct relationship between planned/unplanned and mobile controlled/network controlled handovers.

3. **Forwarding and bi-casting** from the old BS (or another entity in the network such as the "cross over" router) to the new BS. Forwarding is achieved by setting up an IP tunnel between the old BS and the new BS to prevent packet losses due to delays related to the updating of the new location (caused by the Path Updates). This is an integral part of all handover schemes as it is what happens with basic handover solutions embedded in Mobile IP. Bi-casting relates to simultaneous packet delivery to both the old and the new BSs (similar to the *advance registration* feature of MMP explained in section 3.4.3.3, but relating to Handover

Management only). Considering the Bi-casting solution for handovers, they are usually executed from the old BS by instructing it to send packets to the wireless link and to the forwarding tunnel, but can also be achieved from a "cross over" router (Anchors, Mobility Agent...) in the wired network. Again, in some flat network architectures where BSs are not the edge nodes in the network, the "cross over" router can actually be the old BS. The purpose of bi-casting in Handover Management is to reduce packet losses during the transitional phase of the handover. Bi-casting from a "cross over" router inside the network (i.e. not a BS) is not a frequent feature of the handover management proposals since the critical part is route reconfiguration at the old BS. Route reconfiguration (and bi-casting) inside the network, i.e. "cross over" router, can be performed by Path Updates, thus allowing execution of the Handover Management protocol in MHs and BSs only. In addition, if bi-casting is performed by the Handover Management the same feature in Path Updates is excessive.

4. The mechanism for providing executions of other, time critical, IP procedures during handovers. This relates to any **context transfer** between MHs, old BSs and new BSs either through a communication from the old to the new BS or by a direct transfer by the MH (this is generally supported through a generic container and does not affect the message flow of the handover protocol [72]): AAA, addressing information/configuration, link-layer data, QoS data, multicast group membership...Transfer of addressing information is largely dependant on the scenario of deployment. In some planned handover cases, MHs can request a transfer of their identifiers (for example, addresses of Mobility Agents, link layer address, home IP address, *care-of-address*...), which can in turn allow the new BS to complete other mobility functionalities in advance (AAA, Duplicate Address Detection [75], Path Updates/Registrations...).
5. Separation between the mechanisms of the handover protocols and the rest of the mobility protocol. Although this is generally targeted in almost every handover

protocol design, the explicit split between the Handover Management PDI and the rest of PDIs is mostly assumed, rather than specified. A general trend in all handover schemes is to reuse the Mobile IP signalling. Most of the messages specified in Mobile IP for the wireless link are reused with some added extensions. Rather than specifying the operations of the underlying link layer, most schemes merely assume the link layer events (see Handover Decision Making above). The handover schemes also assume that a Path Update protocol handles the location updating procedures beyond the handover stage.

6. Interdependence of handover protocols on the general mobility setup. EMA relies on MANET-based network scenarios and does not require a new *care-of-address* after a handover because in ad-hoc environments, addresses generally do not have a topological significance. Other handover proposal are intended for *Proxy Agent Architectures* and require MHs or new BSs to register with Mobility Agents (i.e. a FA in a particular level in hierarchy), HAs or CHs. Thus there is a direct relation between mobility, i.e. Handover Management, and Address Management, which is avoided in the here-presented design of Handover Management due to the drive for maximum independence of PDI Solutions. Initiation of Path Updates should start after the completion of the handover. This is the case in mobile controlled handovers, i.e. Path Updates are not to be performed until the MH selects a new BS and subsequently connects to it. In a network controlled handover, triggering of Path Updates is sometimes performed before the connection to the new BS is formed (this is an exception rather than a rule).

Handover Design Decisions	Summary
1. Support for Planned and Unplanned handover	Differentiation of procedures when signalling is possible before the change of BSs and when this is possible only after the handover to new BS.
2. Handover decision-making processes	Determining handover actions and network entities, which use knowledge on imminence of handover. Provision of implementation specific link layer assisting features.
3. Forwarding and bi-casting	Recovery procedures for packets potentially lost

	during handovers. Redirection or duplications of packets from chosen network entities.
4. Additional procedures	Context Transfer from old to new BS or any other network entity. Depends on the general connectivity properties (QoS, security...) and interaction with other mobility functions.
5. Modularity of handover management	Relates to the extent of modularity of handover procedures from the rest of mobility functions.
6. Dependency on underlying mobility setup	Relates to whether handover procedures need to be integrated with a specific mobility setup and degree of flexibility in choosing the accompanying mobility features.

Table 6.1. Summary of the Handover Management design decision

This section presented description of different aspects of the research into handover management and associated design strategies (this is summarised in Table 6.1). The next section proposes a new handover protocol with justifications for the protocol features proposed.

6.2 Protocol Proposal for Handover Management[5]

The first design decision when constructing a new handover protocol is to define the extent of the protocol's dependence on underlying link layer technologies. As explained in the previous section, some handover proposals assume link layer scenarios, which can provide significant support for the IP handover and thus facilitate relatively more efficient performances. This support usually comes in the form of a link layer trigger, which indicates to the network layer that a handover is possible, thus expediting the handover decision-making processes. The actual event of a link layer trigger can be utilised in almost all link layer technologies considered (as a step further some handover protocols assume simultaneous connectivity to the old and new BSs). Link layer triggers, regardless of where they occur (in the old BS, MH or the

[5] Note: Design strategy for the handover protocol proposed in this document makes much use of the design decisions used in the "template" proposed in the Generalised IP Handoff framework, which are, in turn, adopted in most of the handover proposals referenced. Since, some of the features of the Handover Management protocol proposed here are specific, design strategy is explained exclusively to highlight the design decisions, which induce the originality of the proposed solution.

new BS), are a consequence of the particular wireless link scenario, i.e. monitoring of wireless link broadcasts during idle periods[6].

The next step in the handover protocol design is determining the degree of handover planning and utilisation of potential link layer triggers in both planned and unplanned handover. The previous section indicated that in the situation where a handover becomes evident (through link layer monitoring) the protocol should aim to achieve the planned handover since it allows an efficient switching of BSs. This comes form the property of planned handovers where a MH is able to prepare for the eventual move to the new BS by requesting the old BS to provide for a temporary flow of packets to the new BS and exchange other relevant information. Additionally, if for some reason the planned handover is not executable, either because of the inadequacy of the link layer, which may not be able to provide relevant triggers, or through a failure of the planned handover protocol steps, a MH must be able to adjust its operation and revert to an unplanned handover.

Thus, a handover protocol should provide methods for performing a planned handover but allow recovery, if the planned handover is not possible, by defining protocol mechanisms for an unplanned handover. This scenario allows the handover protocol to use any underlying link layer support for expediting the performance but at the same time does not assume that this support is always available. Again, an unplanned handover may be the only option for handover scenarios where certain "planning" is not possible, as may be the case with inter-domain handovers. Finally, the handover protocol should be able to accommodate other link layer scenarios. While break-before-make handovers (regarding the link layer connectivity) are generally assumed and do not impede either planned or unplanned handovers, the handover protocol should execute normally even if the make-before-break situation is possible. In this case, the same rules apply: both planned and unplanned handovers are possible

[6] Multi-homed terminals are not considered.

(although if the concurrent connectivity is achievable, planned handover may appear a slightly redundant or at least, an overcautious solution.

The next step in this handover design is determining the choice of the entity controlling the overall handover protocol. This should ultimately be the MH since, unlike the actual network, it knows its application (or user) requirements. A mobile controlled handover additionally alleviates the processing burden inside the network and hence incurs less protocol overhead but as already indicated in Chapter 5 places a significant burden on MHs. A degree of assistance by the network may be beneficial in some circumstances and should be allowed. Thus in a planned, mobile controlled/network assisted handover, the link layer trigger should happen at the MH causing it to run a negotiation procedure with the old BS about the eventual handover to the new BS. Because in some scenarios, there may be more than one target BS, the old BS can assist in suggesting not only one BS, as proposed by the other protocols, but a list of several candidate BSs. However, the MH should eventually make the final choice of the new BS. This may be a highly useful feature since in some setups, the old BS can also be aware of the potential candidate BSs although it is assumed the MH finds out the candidates, usually through link layer monitoring (or via network layer monitoring as in the make-before-break case). The benefit of involving the old BS in the selection of the new BS is that in some scenarios, the old BS may hold extra information about the accessibility of the new BS (apart from the network layer data, BSs can, for example, have link layer data such as Software radio). Possible steps for selection of new BSs can be:

a) All BSs in the vicinity of a MH are the candidates (usually there is only one but scenarios with more than one candidate are possible and should be provided for).

b) MH requests a contact with the new/candidates BSs via the old BS. The old BS can negotiate with the candidates based on its knowledge about the MH's requirements and network policy.

c) The MH makes the final choice of new BS based on its own awareness and suggestions from the old BS. Finally, if more than one BS is chosen by the MH, the old BS can then select a subset of BSs or a single one suitable for the handover.

d) The old BS should receive a notification from the new BS that the handover is possible and relay this confirmation to the MH. This procedure where the MH is informed about the success of the handover attempt through the old BS is more beneficial than if the same was done via the new BS. It enables MHs to receive the notification while being connected to the old BS and thus avoid link establishment delays (for example, if the only communication to the new BS can be achieved through the wireless broadcast channel where the MH needs to fully connect to the new BS and then receive the notification).

e) After the selection and acceptance from the new BS, the MH should send a registration message to the new BS after which a Path Update message can be sent/relayed inside the network to complete the handover.

The exchange of messages between a MH and the new BS via the old BS should achieve the transfer of the MH's identification information (link-layer address, IP address…) to the new BS so that the MH can be accepted based on its already-known credentials. Additionally, other context can be transferred: QoS, header compression information, AAA[7]…

As indicated in the previous section, one of the reasons for providing the additional handover support is to facilitate the installation of a temporary IP tunnel for forwarding/bi-casting packets to the new point-of-attachments (i.e. new BS) while they are still being delivered to the old BS. To prevent packet duplications (which is likely to occur when the same packet stream is transmitted over the wireless mediums

[7] The MH needs to be authorised to use the network and its messages should consequently be authenticated. While this is out of scope of this handover proposal: a way to solve this is to use AAA so that MHs can obtain a session key at the initial access. All the network routers should naturally share a network key, which can be used for encrypting the MH session key and for authenticating messages between routers. A key shared between MH and the new BS can also be used to encrypt traffic over the wireless link.

of the new and the old BS) MHs should be able to detect duplicates and the old BS while the new BSs should not transmit the same packets over their links. Many mechanisms can be provided to prevent packet duplication at the MH; controlled transmission at the old and/or new BS. The old BS could be instructed to buffer packets and not transmit them over its link (once the handover is imminent) and only forward the packets to the new BS through the established IP tunnel. The tunnelling phase must have a short span and should be terminated once the Path Updates are completed. Basically once the Path Updates have been performed the location updates inside the fixed network and packet forwarding from the old to the new BS is simply a protocol overhead, provided all "in flight" packets have been delivered.

Considering the idealised concepts of the Evaluation Framework, it would be highly desirable to design a single generic handover protocol, which would optimally satisfy all efficiency criteria. The main factor inhibiting such a generic scheme from existing is the inevitable functional difference, required even for the handover process, induced by different addressing methods in mobility protocols. Handover Management protocol becomes a mobility sub-protocol only when it becomes fully integrated with the rest of the mobility functions (in this document: when the Handover Management PDI Solution of a mobility protocol becomes integrated with the rest of PDI Solutions, in particular, Path Updates). Most of the handover proposals referenced concentrate on providing the handover support for Mobile IP-based scenarios for Path Updates (i.e. *Proxy Agent Architectures*[8]) and the other PDIs. In such scenarios, an address acquisition procedure during the handover is sometimes necessary and it may differ depending on the scenario. For example, in an IPv6 setup, the old BS and a MH can create a *care-of-address* in a stateless configuration manner and forward it to the new BS. The new BS may additionally perform duplicate address detection (DAD [75]); in

[8] The term Mobile IP-based applies to all protocols using the framework for message transfers defined in the Mobile IP proposal. This does not means that Mobile IP is used for the complete support of mobility since the majority of *Proxy-Agent Architectures* (Regional Registration, Hierarchical Mobile IP...) use the framework only for their control messages.

an IPv4 setup, the new *care-of-address* is usually pre-determined by the address of the new BS, which can perform Foreign/Mobility Agent functionalities. The essential commonality in the approach used for all Mobile IP-based handover management is that the MH obtains a new care-of-address after every handover and that the handover protocol assists in enabling the use of the new *care-of-address* when the MH has handed over[9].

The primary goal of the Handover Management design presented in this document is to make it a sub-protocol of a mobility scenario where the other functionalities are performed by relevant PDIs forming a complete mobility protocol. The particular application of the Handover Management PDI Solution is shown as an integral part of BCMP (see section 5.4.3 and the remained of this chapter) and in the PDI-influenced changes to default MMP shown in section 5.5. The general perspective was that of micro mobility protocols where address change is not bound to handovers which is the property of *Localised Enhanced-Routing Schemes* (see Chapter 2)) and all Packet Forwarding techniques apart from *cascaded tunnelling* (see section 5.3). In such a setup, MHs are not required to obtain a new *care-of-address* during intra-domain handovers and Handover Management is simply a routing remedy until the Path Updates perform the necessary reconfiguration inside the network. While this approach of retaining the (semi) static *care-of-address* during handovers is also used in the EMA handover protocol [50], the general protocol setup is different, because the scheme assumes a MANET-based protocol for the rest of its mobility (additionally EMA requires a temporary address from the new point-of-attachment for "horizontal" Mobile IP signalling). Beside the particularity of control messages and their flows in the specific mobile-controlled/network assisted planned/unplanned handover protocol, this addressing approach presents the major novelty of the Handover Management protocol presented in the following sections.

---

[9] This statement generally describes all Mobile IP-based handover protocols. Although not mentioned, there should be no inhibition in using a Collocated *care-of-address* in the IPv4 proposal, which makes the statement only partially correct.

6.2.1 Planned Intra-domain Handover

This section explains the basic protocol features of planned handover. Detailed flow of all control messages can be found in one realisation of this framework shown in Appendix 1. Figure 6.2 shows the principle operations of planned handover.

A link layer trigger can happen both at the MH or the old BS. Wherever, it leads to the negotiation of candidates ultimately causing the MH to request a handover to one or more candidate BSs. The MH sends its request to the old BS, which forwards the request to the new/candidate BS(s). At this stage, the old BS can relay any relevant context about the MH to the new BS. The old BS then informs the MH about BSs that have positively acknowledged the handover request. MH uses this indication and registers with the new BS and the old BS starts tunnelling packets to it. The new BS performs Path Updates and confirms the MH's registration. After the Path Updates have reconfigured the routing entries so that the traffic is diverted to the new BS, packet forwarding at the old BS should be terminated. This can be achieved explicitly by a control message from the new BS or another routing entity in the mobility setup, possibly a "cross-over" router. Alternatively the tunnelling can be terminated by a timeout at the old BS. Models for forwarding tunnel termination are given in the adaptation of the Handover Management protocol for MMP and BRAIN Mobility Compromise Protocol in section 6.3.

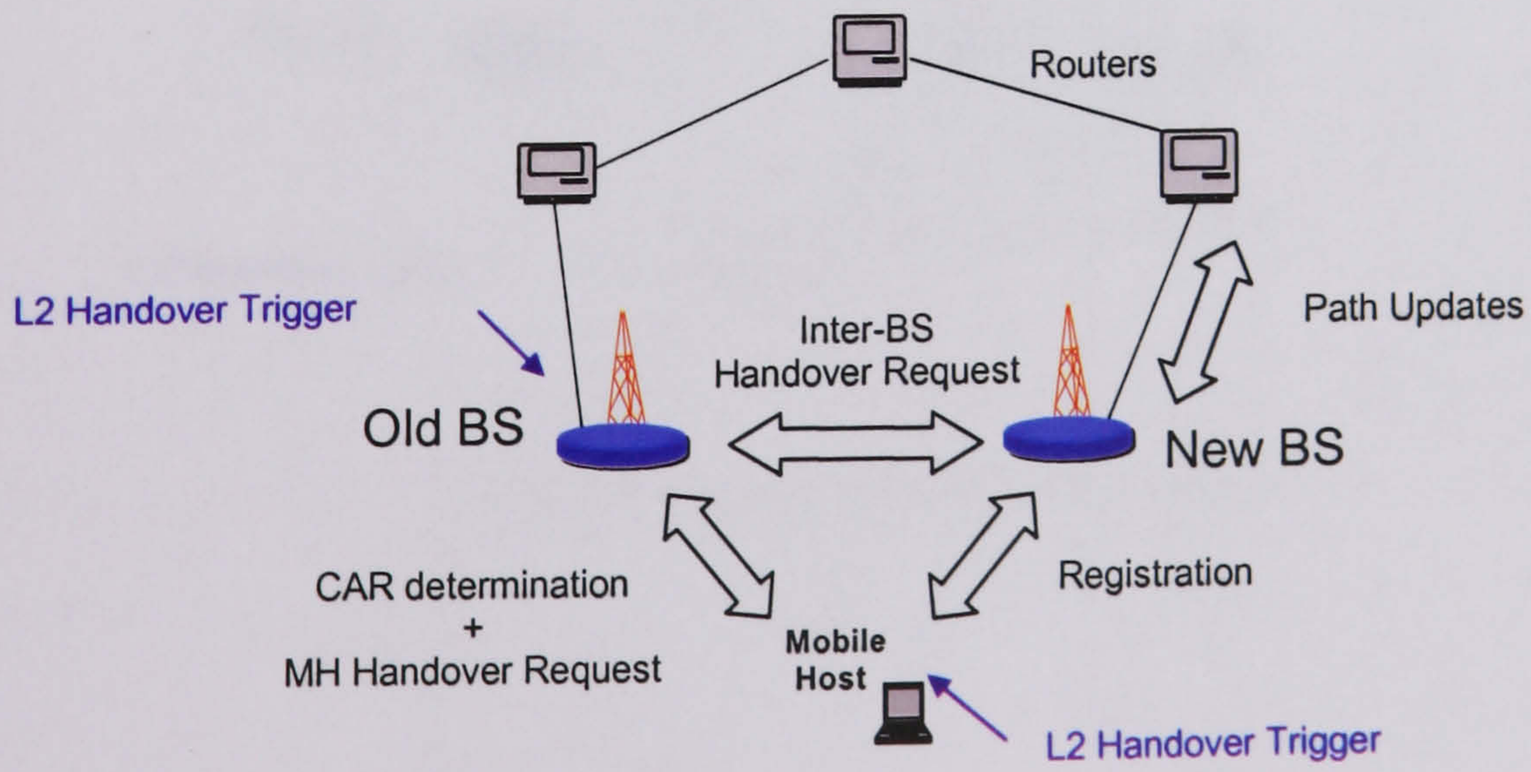


Figure 6.2. Planned Handover Concept

6.2.2 Unplanned Intra-domain Handover

The principles of the unplanned handover are shown is

Figure 6.3. Detailed flow of all control messages can be found in one realisation of this framework shown in Appendix 1.

MH receives an indication that an unplanned handover is needed (a link layer trigger such as a loss of connection to the old BS), it establishes a connection with the new BS and registers to it, conveying the identification of the old BS in the registration request message. The new BS contacts the old BS and negotiates the necessary context transfer along with the establishment of the IP tunnel for forwarding packets from the old BS. The new BS performs Path Updates and acknowledges the registration from the MH (This last step could be done in parallel with the global handover processes of an intra-domain handover, provided that the new BS already has a security context for the MH).

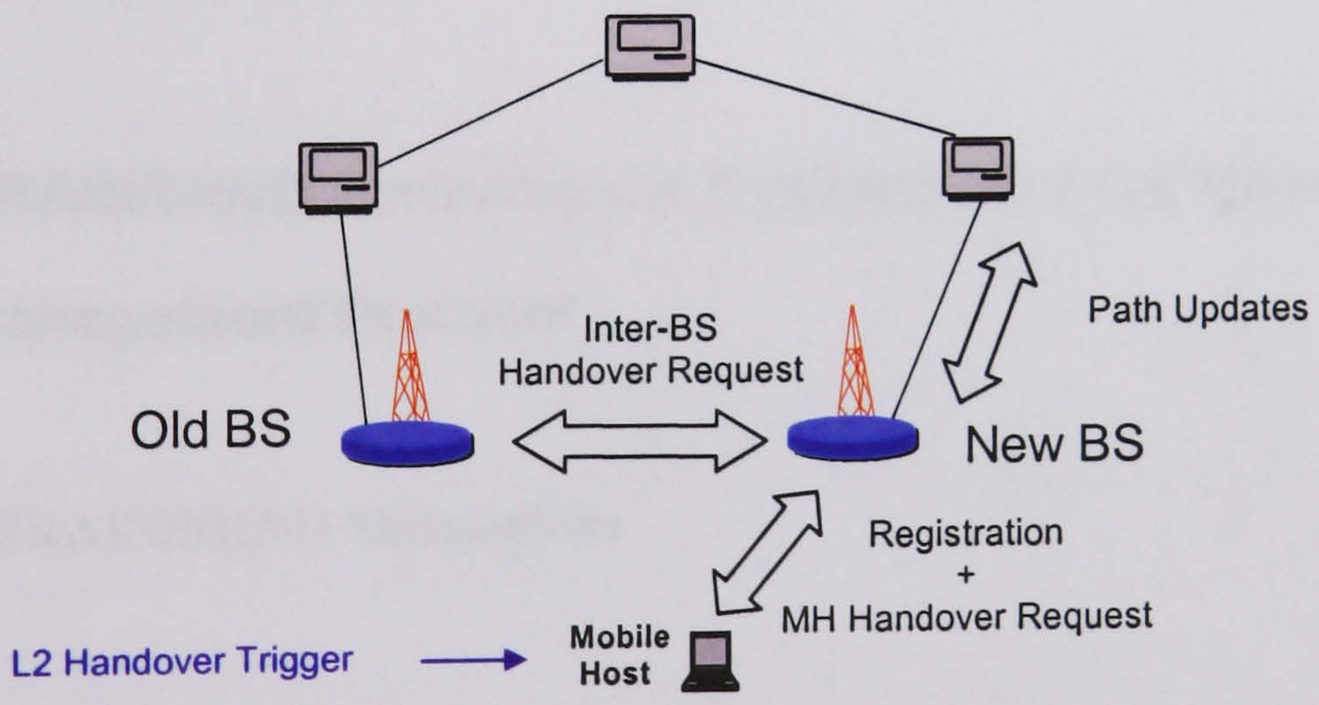


Figure 6.3. Unplanned Handover Concept

6.3 Integration of the Handover Management Protocol Design Issue with the rest of Mobility Protocol

As mentioned already, the Handover Management framework designed in the previous section is intended for integration with the rest of the PDI Solutions, which collectively form a micro mobility solution similar to the setup of the *Localised Enhanced-Routing Schemes* and the three Packet Forwarding techniques excluding the *cascaded tunnelling*. The distinguishing element of the handover operation is the assumption that the handover protocol is not involved in the address acquisition and assumes a static/semi-static *care-of-address* for the duration of the connection in a network, or in more precise terms, during intra-domain handovers.

In following sections, one realisation of the handover protocol framework is shown, when integrated with BCMP, by observing the simulation results obtained in the BRAIN/MIND projects. This is used for drawing general conclusions on the performance of the Handover Management PDI. Integration with other protocols such as MMP is also discussed.

6.3.1 BRAIN/MIND Performance Evaluation of the Handover Management Protocol

6.3.1.1 BRAIN/MIND Simulations

The following text presents performance evaluation of the Handover Management PDI in BCMP. The results shown are taken from the public results of simulations and performance analysis included in the BRAIN[10] and MIND projects public deliverables

[10] As explained in section 1.6 and Appendix 2, the here-contained results are mainly related to the results of research in BRAIN project but have also been performed in the MIND project being its direct

[47][89], which contain a more extensive set of performance evaluation and related issues. Testing of protocol mechanisms of BCMP was performed via simulations using ns-2 (see Appendix 4 for description of ns-2 simulation tool and the reasons for choosing the particular simulations tool in the projects). **The results are analysed with respect to the work presented in this and previous chapter but have been performed as a join effort inside the projects and included as their final results. Hence, the simulations strategy and the results presented are attributed to the projects and not to the author (this is further explained in section 1.6 regarding the author's contribution in the projects).**

As pointed out, Hanover Management protocol is intended for integration in a mobility protocol by combining with other PDI Solutions. One such protocol is BCMP explained in section 5.4.3 (more details are in [47] [61]) developed in the BRAIN project. Intended testing of the Handover Management protocol can only be performed as a part of the complete mobility protocol. This is because the main objective of the simulation is to extract performances during handovers for which packet flows are required to MHs using the Handover Management protocol and this inevitably requires the whole mobility protocol to be in place. However, the performance of the whole protocol relative to handovers is significantly dependent on Handover Management protocol since it largely affects packets losses and latencies especially for the planned handover case. Thus, the Handover Management protocol designed in this chapter is tested as a part of BCMP but its performance can be related to any integration scenarios such as MMP explained in section 6.3.1.2 and already noted as the initial design objective for Handover Management PDI explained in section 6.1 and [47]. Focus on general Handover Management performance properties is maintained throughout the presentation and explanation of BRAIN/MIND

follow-up. However, the work does not cover some specific MIND extension to mobility solution as its scope of research was broadened.

simulation results for Handover Management PDI and is used for deriving generic conclusions on Handover Management performances.

If an observation is made disregarding the existence of Handover Management protocol, BCMP performance (especially Path Updates) is not as efficient concerning the handover performance as some *Localised Enhanced-Routing Schemes*. This is because of the placements of the serving ANP thus *handover distances* may not be minimal for every handover as in Cellular IP, HAWAII and MMP. However, an extremely dense population of ANPs may still provide this requirement. As an example of this property Figure 6.4 shows conceptual representation of Handover Management and Path Update effects on packet flows and resulting disruptions during handovers. Router as shown in the figure contains the host routing entry in the network, which needs to be updated with the new location of MH. In MMP this is "cross over" router as explained in Chapter 3 and 4. Its location is dynamic and depends on the closest common router for the old and new routing tree to old and new BS. However, in BCMP this dynamic property in not present and the host routing entry is always contained in ANP which is fixed (assuming MH does not change its serving ANP since this change is not tied with the mobility of MH). This observation can be used for analysing general features of Handover Management protocols in other mobility protocols such as MMP.

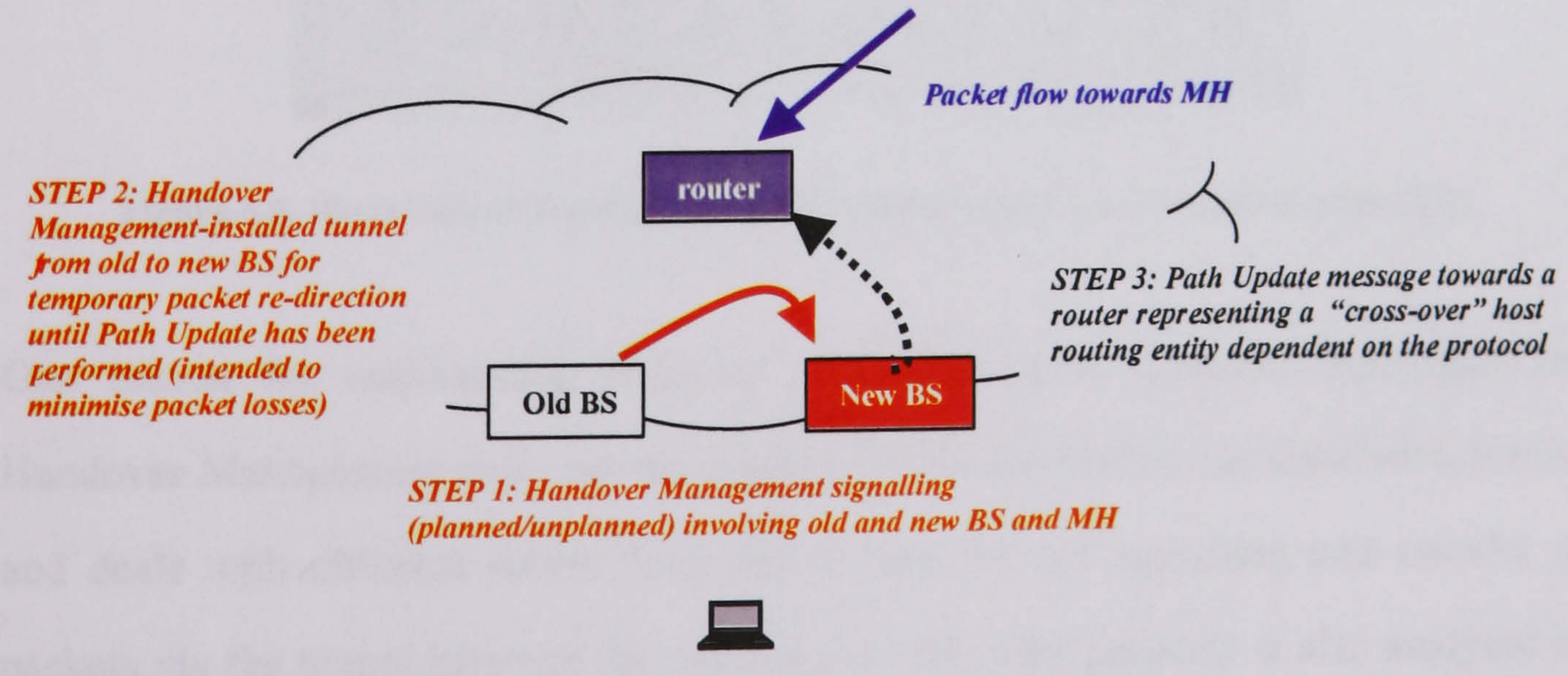


Figure 6.4. Conceptual differentiation between Handover Management and Path Updates effect on packet flows for arbitrary mobility setup

The actual network topology cannot be predetermined and varies according to the deployment scenarios. The focus of this study is on the hierarchical topology as shown in Figure 6.5 and the **partial mesh** topology suggesting that all the links are not interconnected as a full mesh topology would assume as shown in Figure 6.6. A hierarchical topology was already introduced in the simulations of MMP in Chapter 4, where the key property is that all routers have a single upstream neighbour converging at the network gateway. Unlike hierarchical topologies, mesh topologies generally do not have any rules of interconnection between routers. The interconnections are usually implemented according to the operator's preference. Full mesh refers to the case where all routers are mutually connected. Partial mesh is a more realistic scenario than full mesh with a high level of mutual interconnection of routers and BSs.

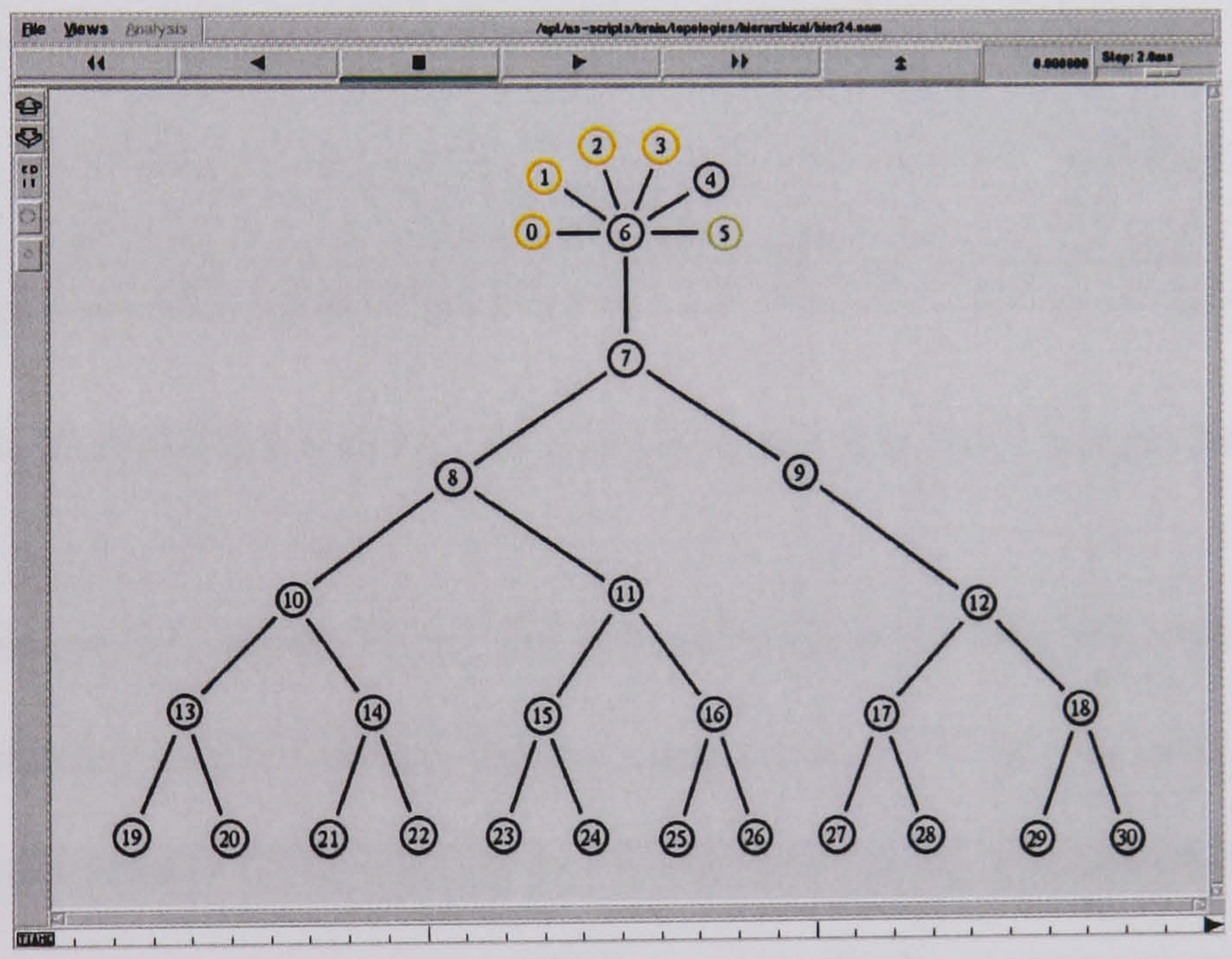


Figure 6.5. Hierarchical Topology used in the simulations (ns-2 image) (source [89])

One reason for emphasising different performances in different topologies for Handover Management case, can be extracted from the analysis of simulation results and deals with different routes from old to new BS for signalling and transfer of packets via the tunnel between the old and new BS. This property is also analysed in section 5.2.2.3. Regarding mobility protocols which rely on Path Updates for

handover transfer from old to new BS, as the case with MMP shown in chapter 4, dependency on topology is not a critical factor in examining the impact of the protocol on *handover latencies*. This follows the assumption that Path Updates are addressed to the gateway and that routing in the network is shortest path. This can apply to MMP and some other protocols such as Cellular IP.

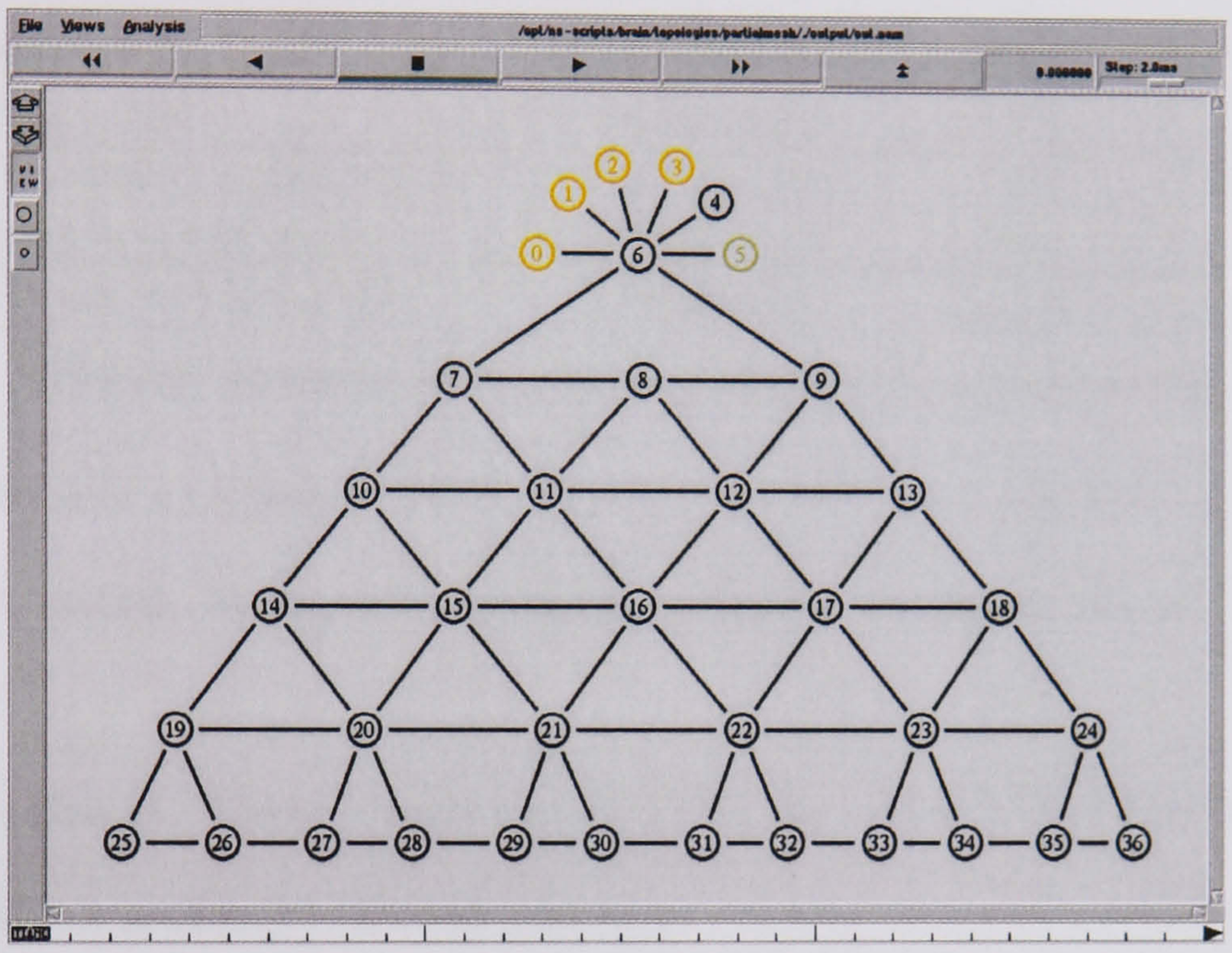


Figure 6.6. Partial Mesh topology used in the simulations (ns-2 image) (source [89])

As shown in Figure 6.5 nodes '7' to '30' are the different network routers. Node '7' is the network gateway to the Global Internet and nodes '19' to '30' are BSs. Nodes '0' to '5' represent various CHs and node '6' in this study represents the MHs' HA, which forwards packets to the current address of the MH. For the partial mesh case as shown in Figure 6.6 nodes '7' to '36' are the various network routers. This network topology has 12 BSs and can also support two network gateways to the core network, nodes '7' and '9'. Nodes '25' to '36' are the BSs. Nodes '0' to '6' are the identical as in the case of hierarchical topology.

Regarding the physical characteristics of the network, Table 6.2 summarises the different values used. The values are the same for both network topologies. The

bandwidth of the links is relatively high compared to the parameters chosen for the simulations of MMP (see section 4.1.2), which are also similar to the referenced external simulation attempts. The network is also overloaded with traffic for a large population of MHs not only MH used for extracting the simulation results. The particular reasons for this simulation strategy and the entire scope of the simulations can be found in [47][89].

Link	Bandwidth	Delay
CH(s) – HA	50 Mbytes/s	5 *ms*
HA – network gateway	100 Mbytes/s	5 *ms*
Network wired links	20 Mbytes/s	1 *ms*
Network wireless links	5 Mbytes/s	0.5 *ms*

Table 6.2. Simulated network physical characteristics (source [89])

Regarding the tree and partial mesh topology, number and position of ANPs in the network was varied. In the tree topology different configuration of ANPs are as follows:

- **Configuration A:** There is only one ANP in the network and this is node '7' in Figure 6.5, which is also the network gateway.
- **Configuration B:** There are two ANPs on the network: nodes '8' and '9'.
- **Configuration C:** There are three ANPs: nodes '10', '11', '12'.
- **Configuration D:** Finally there are six ANPs: nodes '13', '14', '15','16', '17'and '18'.

In the partial mesh topology configuration of ANP are as follows:

- **Configuration A:** There are two ANPs in the network, which are nodes '7' and '9' in Figure 6.6, which also act as network gateways.
- **Configuration B:** There are two ANPs on the network: nodes '10' and '13'.
- **Configuration C:** There are three ANPs: nodes '14', '16', '18'.
- **Configuration D:** Finally there are again three ANPs: '20', '22', '24'.

Different configurations of ANPs mainly relate to the performance of Path Updates in BCMP but also show the interactions with the Handover Management protocol from

which conclusion can be made on the performance of the Handover Management when integrated with other PDI Solutions set such as MMP (see Figure 6.4).

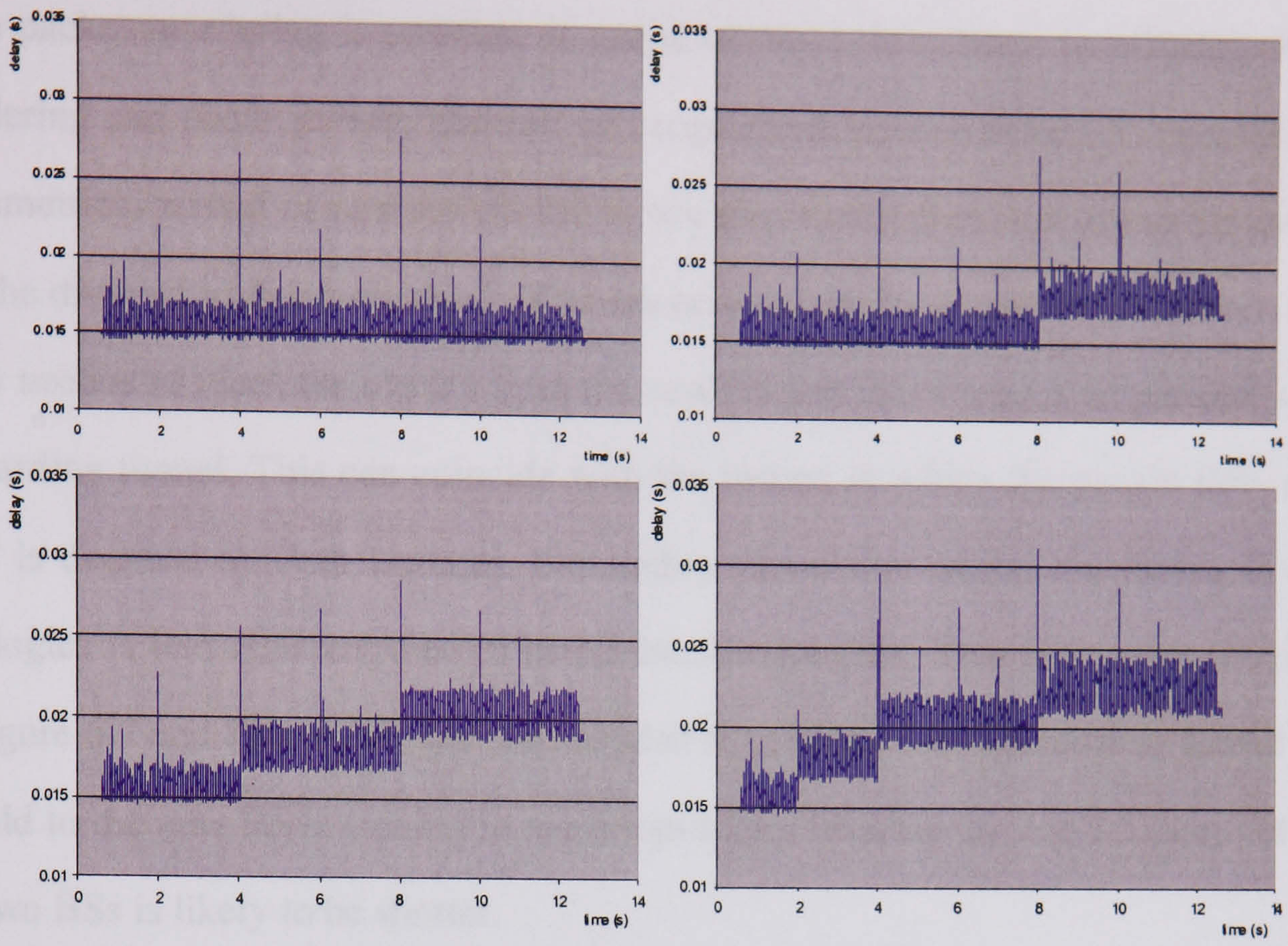
Simulations results present behaviours of packet transfers and delays of messages used in the handover execution. Network is heavily loaded with traffic by having a large population of MHs (24 are constantly attached to the network) with various traffic streams for uplink and downlink traffic using UDP and TCP transport protocols. MHs register at the start of simulations with the ANP closest to their cell. Results are extracted for 2 moving MHs, which move in a straight line from one end of the network to the other performing 11 equally timed handovers (other MHs are stationary). Cell coverage overlaps thus graceful physical handover is assumed as in the case of MMP simulations in Chapter 4. As mentioned in the protocol specifications (this chapter and Appendix 1) some features of the Handover Management protocol are implementations specific. During the planned handover MH, when receiving the handover preparation acknowledgement message (Host Handover Reply) from the current BS, immediately performs the link layer handover and sends the handover message (Registration Request) to the new BS thus avoiding any duplicate packets. When receiving the Path Updates message from the new BAR ANP may also exercise control regarding when to update the path to the new BAR. This operation in the simulations is assumed to happen immediately upon receipt of the Path Update message. Various multimedia traffic sources are used in simulations both for uplink and downlink with various traffic types of average throughput of 2.5Mbits/second and overall packet size of 60 bytes (IPv4) (more on traffic models used can be found in [47][89]). The following explains basic properties of the Handover Management from the performance evaluation conducted in the BRAIN/MIND projects:

- Planned Handover: The first expectation from the performance of planned handover is complete elimination of packet losses. This should come as a consequence of the temporary redirection of packets from old to new BS during the handover

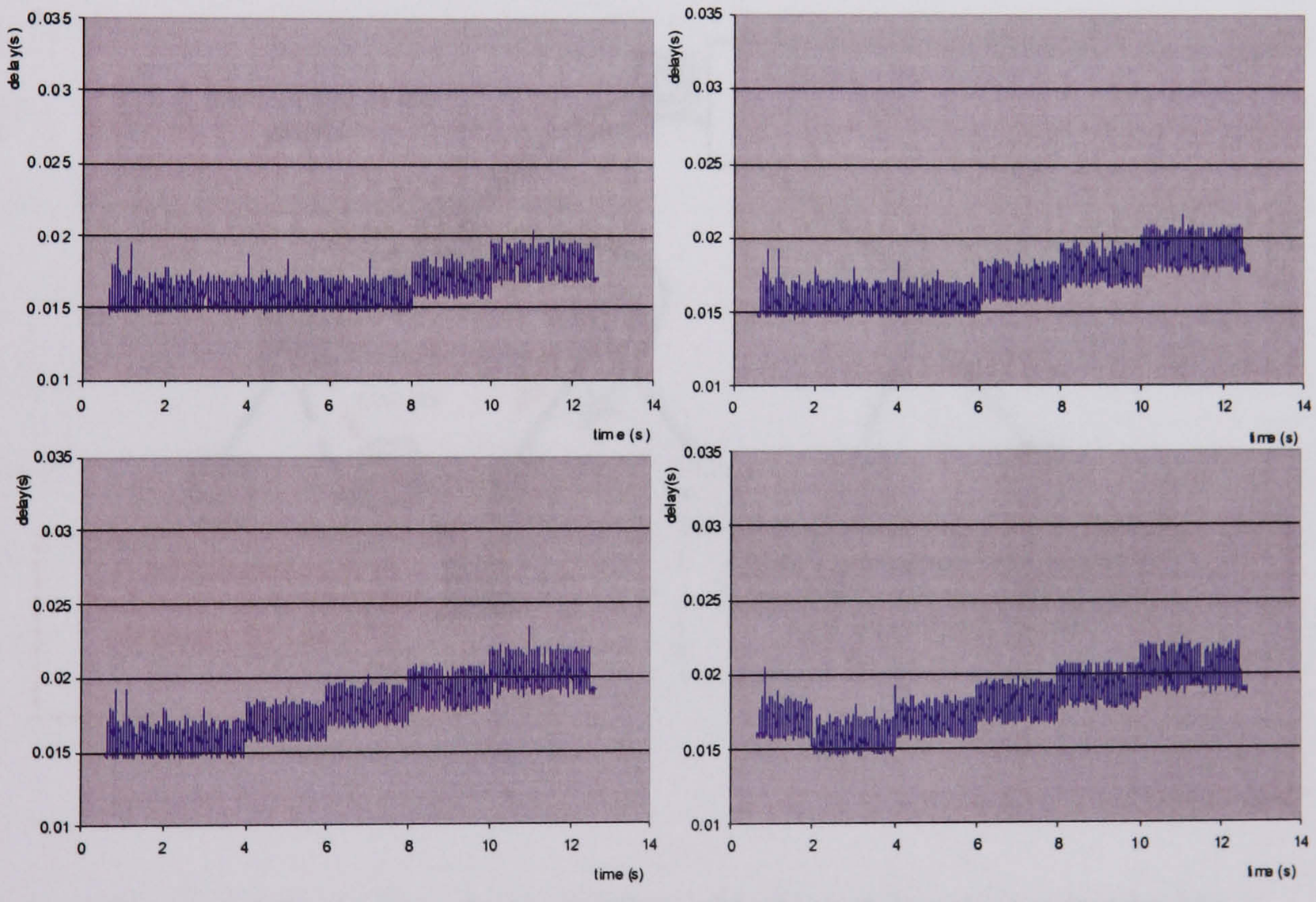
preparation stage. Hence no packets are lost while MH handovers between BSs. This property is noted in performance evaluation of the planned handover shown in BRAIN/MIND results in [47][61][89]. Due to the fact that packets are sent to the new BS from the old BS during the completion of the Path Update process there is a possibility that some packets may arrive out of order. In BCMP, Path Update process relates to updates of serving ANP. This is because packets earlier in the sequence of packets sent to MH, are sent from the old BS back to the new BS while other consecutive packets are redirected from the ANP once Path Updates are completed. Hence, packets redirected from ANP can reach the new BS before all or some of the packets sent using the temporary tunnel from old BS. The reordering of packets in planned handover is examined in more detail in BRAIN simulations shown in [61]. This extent of the reordering is dependent on the network topologies as depicted in Figure 6.7 and Figure 6.8 exemplifying processes of packet flows in hierarchical and partial mesh topologies respectively. Two basic properties of the planned handover can be further highlighted from the performance evaluation conducted in the BRAIN/MIND projects and are concerned with general dependency on Path Updates and network topology. Graph 6.1 and Graph 6.2 show UDP traffic packet delays during handovers for the whole set of handovers performed by MH for hierarchical and partial mesh topology respectively. The first conclusion is that in partial mesh topology, these delays are smaller due to the fact that packets, tunnelled from old BS to new BS, reach the new BS using a shorter route than in hierarchical topology where they need to travel "back" to the common router for two downlink routes for old and new BS. In addition, due to the shorter route in partial mesh topology, there are no peaks in packet delays occurring at handover times in hierarchical topology. Another dependency on packet delays and topologies can be observed considering different configurations A, B, C and D for placements of ANPs in both topologies, which indicate dependency on Path Updates. In configuration A for both topologies, the serving ANP is placed at the "top" of the network closer to the gateways. Due to this,

delays are very similar for the same configuration since paths to the initial and the rightmost BSs at the end of the simulations have similar hop distances. However, as configurations are changed and ANPs are closer to the BSs, topologies play a significant part in packet delays as evident from the two figures for both topologies. If configurations D are taken as an example of this statement Graph 6.1 shows that in hierarchical topologies packet delays are increased compared to the same configuration case in partial mesh topologies shown in Graph 6.2. Again, this property comes due to the shorter hop distance in partial mesh topologies, which are emphasised in configuration D since ANPs are located closer to BSs. Hence, when MH performs handovers away from the initial BS, delays in hierarchical topologies are emphasised due to larger hop distance for packets travelling to current BS of MH (e.g. in hierarchical topology configuration D when MH's serving ANP is node '13' in Figure 6.5, and MH is currently attached to BN '30', there are 7 hops for packets sent to MH from its serving ANP).

- Unplanned Handover: Unlike the planned handover, in the case of the unplanned handover, latency associated with changes of points-of-attachment is inevitable. This is due to the fact that MH installs the temporary tunnel from old BS to new BS **after** it has connected to the new BS (see section 6.2.2). This also coincides with the process of performing Path Updates from new BS. Packet losses associated with unplanned handover latencies have been noted in BRAIN/MIND performance evaluations showed in [47][61][89] for applied network characteristics and traffic rates. Regarding the properties observed in the case of planned handover these follow the same pattern in the case of unplanned handover since the same topologies and configurations of ANP was applied (see section A4.2 in Appendix 4). One particular property of unplanned handover is increase in packet delays during handover times since "no planning" is applied and packet losses are registered due to the need for updating the old BS and ANP after handover to the new BS.



Graph 6.1. Planned Handover: Packet End-to-end Delays – Hierarchical Topology Configurations A, B, C, D (source [89])



Graph 6.2. Planned Handover: Packet End-to-end Delays – Partial Mesh Topology Configurations A, B, C, D (source [89])

The simulations of both planned and unplanned Handover Management protocols in BCMP shows that, depending on the topology where the protocol is being deployed,

some packet reordering is possible as noted in simulations show in [47][61]. Packet reordering can occur in both planned and unplanned handovers and is caused by the asymmetrical arrival of packets via the forwarding tunnel from the old to the new BS and the diverted traffic at the ANP. The essential reason for packet duplication is the delay needed to reach the old BS from the new BS and the reception of packets via the forwarding tunnel. This can coincide with the instant at which the packet flow at the ANP is diverted by Path Updates. Simulation shows that packet reordering in mesh topologies is less apparent than in hierarchical topologies. This is because, as shown in Figure 6.7 and Figure 6.8, the time needed to configure the forwarding tunnel from the old to the new BS is smaller in mesh topologies because the hop distance between the two BSs is likely to be shorter.

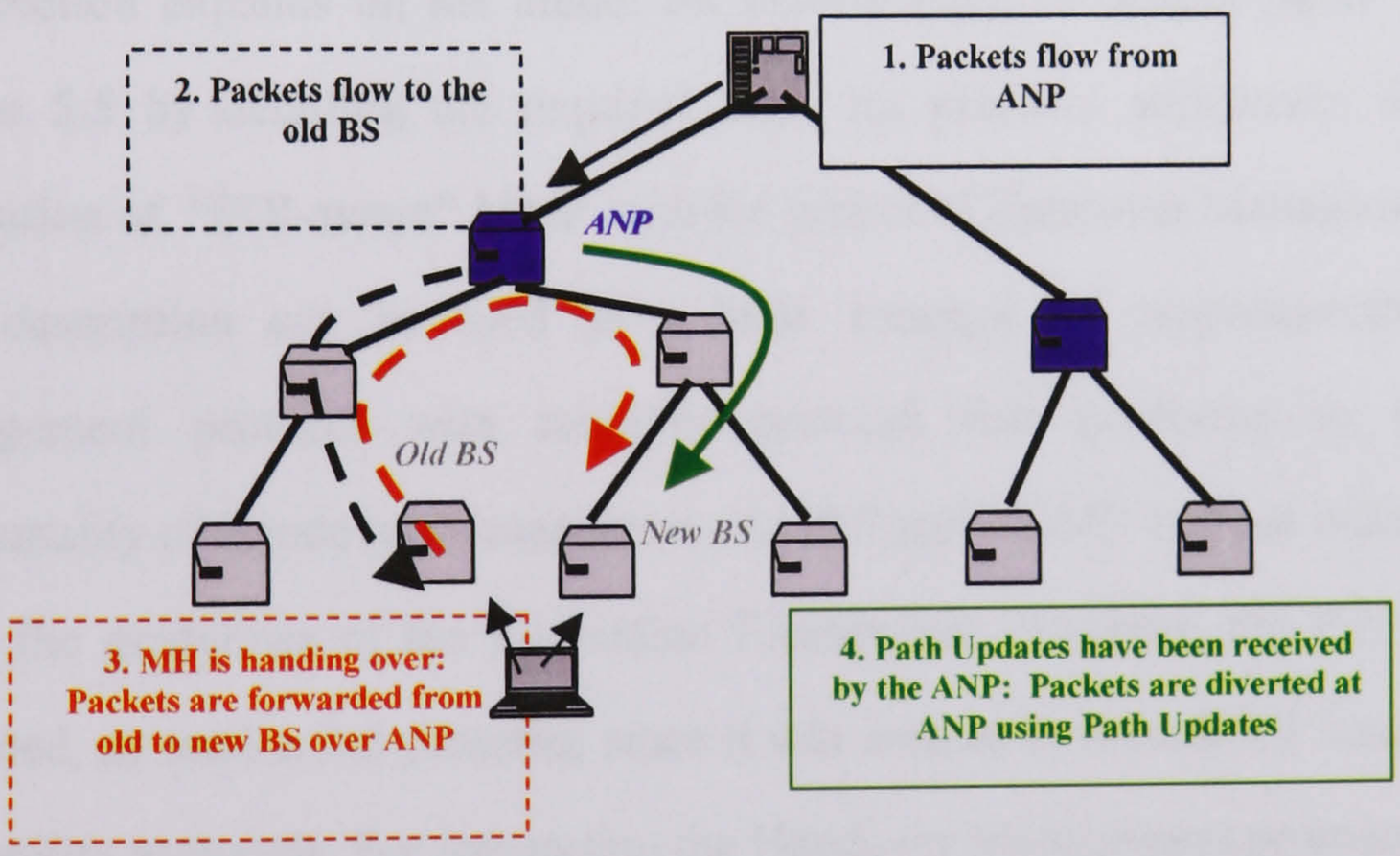


Figure 6.7. Packet reordering in hierarchical topologies during handovers

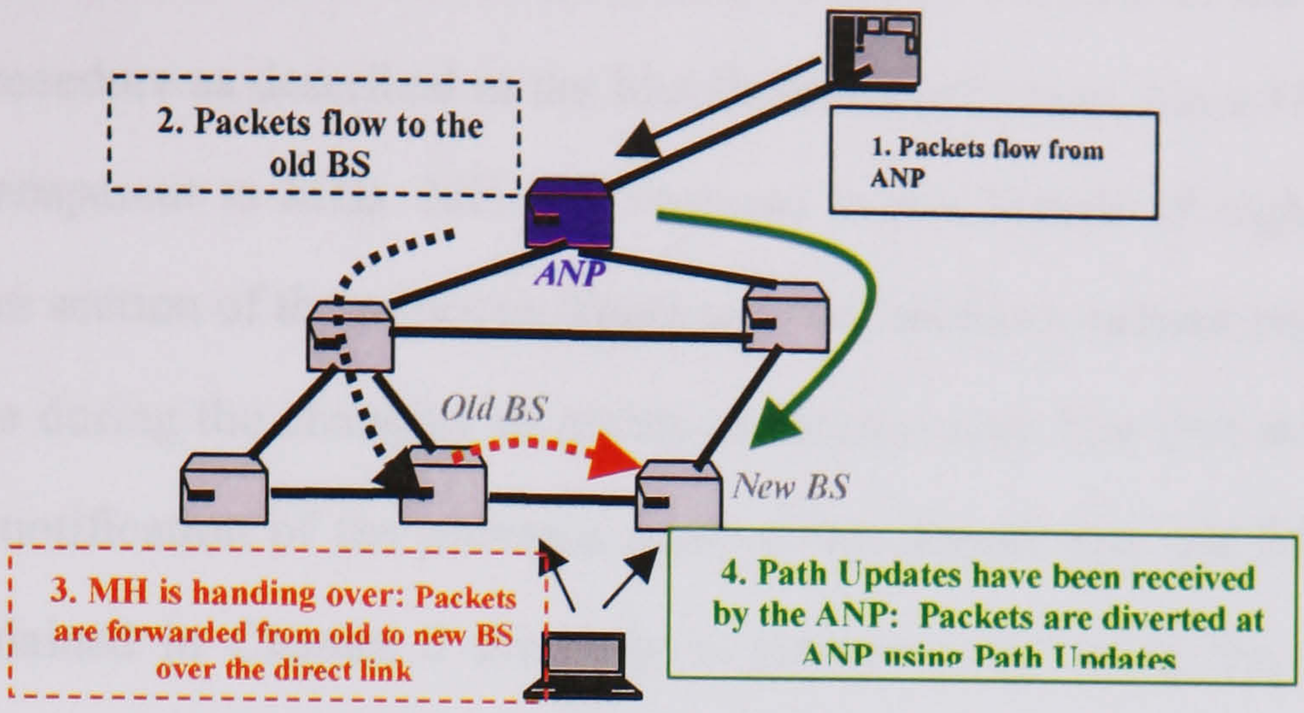


Figure 6.8. Packet reordering in mesh topologies during handovers

6.3.1.2 Integration methods of the Handover Management protocol and MMP

This section expands on the model for enhancement of default MMP presented in section 5.5 by detailing the required steps for protocol adjustment for achieving integration of "PDI-aware" MMP with the proposed Handover Management protocol. This description can be used as a basic example of supplementing Handover Management protocol with mobility protocol that conforms to the required functionality of Handover Management and PDI split. MMP was not initially designed along the guidelines of the Evaluation Framework. However, the PDI split can be extracted, as section 5.5 presents, since it was created to include all functionalities of IP mobility protocols. For integrating the Handover Management protocol with MMP, procedure consists of replacing features of MMP, which form default handover management PDI with the new Handover Management sub-protocol. In MMP, Handover Management procedures are not complex and involve simple change of point-of-attachment followed by triggering of Path Updates. Thus, potential introduction of the Handover Management protocol in MMP is more addition of the new sub-protocol rather than replacement of existing and less efficient mechanisms.

As explained in Chapter 3, the handover procedure in MMP consists of the network layer detection procedure as described in the Mobile IP specification, since MMP was intended to be transparent to MHs. MHs are required to use Mobile IP signalling in the *wireless access* section of the protocol. There are two handover related steps of the MMP's procedure during the changing of points-of-attachments. The first step of the procedure is the notification of the previous point-of-attachment (i.e. old BS) to the new BS. As explained in Chapter 3 this step is performed by using the Previous Foreign Agent Extension during the registration to a new BS. The extension carries the address of the old BS to the new BS. The second step of the procedure is the creation and transmission of the *MMP Instruct* message to the old BS, this being enabled by the information supplied in the first step. The *MMP Instruct* message is used to start removal of the old routing tree branch from the old BS, or in other words, to trigger a "negative" Path Update message. MMP does not have a feature, which is common to all other handover management approaches: forwarding/bi-casting from the old BS to the new BS. Bi-casting is possible, but as a Path Update-facilitated feature from a "cross over" router described in the *advance registration* feature. Thus, integration of the handover management protocol with MMP requires alteration of the routing functions in MHs and BSs to allow them to perform all the steps of the planned and unplanned handovers. Modifications are not required inside the wired network. These modifications of MMP needed to support the integrated handover management protocol are:

- MHs and BSs need to deploy the Handover Management protocol as designed in the previous sections. Hence, all default features of MMP, which correspond to default handover procedures, become obsolete.
- The handover management procedure includes informing the new BS about the address of the old BS. Hence, the new BS does not expect the Previous Foreign Agent Extension.

- Since in the proposed Handover Management PDI, forwarding of packets occurs from the old BS, the new BS is not required to transmit a *MMP Instruct* message as soon as MH hands over to it. By integrating the Handover Management PDI with the rest of the functions of MMP, the old routing tree branch is not immediately removed after the handover as specified in the default MMP. In fact, the old tree branch is still required for packet delivery to the old BS to facilitate the use of the temporary forwarding tunnel from the old to the new BS.
- When interpreting MMP operations with respect to the PDI split of the Evaluation Framework, Path Update PDI is essentially represented by the CBT Join Request and CBT Join Reply messages of MMP (additionally, the Registration Request and Reply). Triggering of transmission of Path Updates should be in accordance with the handovers specifications in the previous section.
- As indicated in the description of the Handover Management, forwarding from the old to the new BS should have a short span and should terminate as soon as the Path Updates have diverted the packet flow to the new BS (packet flow is diverted at the "cross over" router). The Handover Management framework does not include specification of an exact mechanism for terminating the forwarding from the old BS as this can be specific to the actual mobility protocol. As indicated in the specifications of handover, solutions could include timeouts at the old BS or alternatively an instruction to the new BS to send a control message to the old BS and hence cause termination of the forwarding. A *MMP Instruct* message could be re-used as a solution since it is already a part of MMP. The new BS would simply transmit a *MMP Instruct* to the old BS. When the old BS receives the indication that the forwarding to the new BS is not needed, it should remove the tunnel and remove the old routing tree branch using the CBT mechanisms defined in the MMP specifications.

CHAPTER SEVEN

Conclusions

7.1 Contribution of the Thesis

This thesis addresses development of mobility solutions in the Internet. The need for supporting wireless and mobile access in the Internet is increasing with the expansion of laptops and hand-held devices and wide adoption of IP technology. The work presented is predominantly directed towards achieving mechanisms and models for improving mobility support in the IP network layer. Although IP is the universal network layer for many different telecommunications environments, the results of the work presented in the thesis are directly applicable to the specific Internet Protocol network scenarios mentioned in section 1.4. This claim is mostly influenced by the performance testing parameters and analysis of developed protocols shown in Chapter 4 and Chapter 6. As mentioned in section 1.4, research presented in the thesis aims at maintaining a generic approach to development of IP network layer functionality. Specific implementation scenarios (where applicable) are interpreted as possible design decisions rather than explicit quantitative input parameters (e.g. design

decisions in Chapter 6). Analysis of developed protocol features often attempts to include all generic issues of the protocol's operation hence giving estimates about its performance in any network scenario. One example of this is the additional mathematical performance analysis shown in Chapter 4, which extends network topology cases and mobile host populations from the ones applied in the simulations. Analysis of simulation results shown in Chapter 6 is focused on explaining general properties of the planned and unplanned handovers in various cases of network topology and remaining mobility functionality.

This thesis applies a specific approach in describing development of IP mobility protocols. Research results presented are often descriptive and textual in nature rather than being similar to the manner in which protocols and their operations are typically described in the standardisation bodies (e.g. IETF protocol specifications...). It is believed that the approach of this thesis is beneficial since it presents an attempt to highlight concepts extracted from protocol functionalities thus offering a reader a chance to become acquainted with the design theory of IP mobility protocols rather than a discrete set of functions of a particular protocol.

In order to start the consideration of particular mobility solutions, Chapter 2 offers a detailed explanation of different concepts of mobility in the Internet. This is then used as the start for explanation of Mobile IP being the reference mobility protocol in the Internet. Introduction of the abstract mobility model is believed to present a high-level understanding of which IP deficiencies mobility solutions are attempting to overcome and provides an understanding for the subsequent classification of IP mobility protocols.

This knowledge is then applied in Chapter 3 by showing all steps of the design of the here-proposed MMP. The work presented in relation to MMP starts with the thinking process behind its creation. Inclusion of the alternative routing methods, that is, *sparse* mode IP multicast as the key feature of MMP, is considered to be the most crucial property of the protocol. The presented analysis of mobility with IP multicast and

similar solutions in the Internet, reveals that the solution adopted presents a more efficient model for fast and easy deployment of mobility support. MMP is then carefully developed as a fully functioning protocol where all the operational details are considered. Naturally, the protocol presented is the most recent version of the protocol, which was preceded by some trial and error versions. As already indicated, some of the operations of the protocol could be performed without the integrated IP multicast where all control messages perform the same actions but as independent features of the mobility protocol. This was deliberately avoided because of the stated emphasis on IP multicast and the intention of showing integration of seemingly different mechanisms, that is, mobility and multicast.

Chapter 4 offers a comprehensive insight into examination of performances of IP mobility protocols by devising simulation strategy and presenting results for MMP (being the focus), Mobile IP and Hierarchical Mobile IP, which are used to highlight different performance properties shown in the simulation results. The results reveal inevitable trade-offs associated with performances of IP mobility protocols. In the highlighted case of MMP, this is represented in the set of results for handover performance and protocol overhead where the overall conclusion is that improvement of one aspect typically assumes degradation of the other. The simulation results are further supported with introduction of mathematical models for validating the obtained results and their subsequent expansion for additional performance analysis of handover and protocol overhead for some extra scenarios of mobile host populations and network topology.

Research into mobility protocols and particular development and testing of MMP reveals the trade-offs in performances of IP mobility protocols and the, sometimes subjective, nature of evaluation of their mechanisms. This provides the basis for approaching the problem from both generic and practical stance and is used as a starting point for the Generic Mobility Design Model as presented in Chapter 5. The problem of IP mobility development is initially broken down into abstract research

disciplines of the Evaluation Framework and it is then followed by its practical application thus constituting the model. The essence of the model is the idealised split of IP mobility protocol functionalities into Protocol Design Issues and their resulting Solutions. Chapter 5 then describes how the modularity offered by the PDI split can be used for managing designs of Mobility Management solutions with the intention not being on creating a single final solution but on a model for achieving synchronisation of particular design principles and deployment requirements and using them as constructive input for "moulding" desirable mobility solutions either for creating new mobility protocols or for modification of functionalities of existing protocols. This is exemplified by describing the design processes that lead to the development of mobility solution in BRAIN (and MIND) projects and by showing possible enhancements of MMP using the PDI split. The overall aim of Chapter 5 is to describe the instrument by which the Generic Mobility Design Model can be applied practically for the development of mobility solutions. Chapter 6 continues the work explained in Chapter 5 by showing the design, thus realisation, of a particular element of the PDI split of the Generic Mobility Design Model: the Handover Management protocol (PDI Solution). The protocol is developed as a generic handover solution for any mobility protocols that conforms to the PDI split. This is exemplified in the functionality of BCMP and enhancements to MMP shown in Chapter 6 and can be further applied to other mobility solutions. Specific design processes shown in Chapter 6 are intended to highlight the research and decision making that preceded the specification of the mobile-controlled/network-assisted planned and unplanned Handover Management protocol. It is believed that different design preferences can be used to apply the same methodology as shown in Chapter 6 and achieve a different Handover Management protocol. Handover Management protocol is tested as part of BCMP (author contributions in BRAIN/MIND projects are described in section 1.6) but the simulations are configured to extract performance parameters relative to the

developed Handover Management protocol to assist in interpreting its general performances.

As noted in the Motivation for the Research (Section 1.1), the research results presented in the thesis are intended to offer easy realisation in considered deployment environments. It is believed that this is achieved with MMP which is based on IP multicast and can also resort to standard Mobile IP, should the network operator decide not to deploy any further mobility support. The Generic Mobility Design Model is intended as a tool for constructing mobility protocols in chosen deployment environments.

7.2 Future Research Directions

The work presented in this thesis opens the way for further research in various directions:

- **Further mathematical validation and modelling of performances of mobility protocols:** An observation made in the previous section is that the research presented in the thesis mainly applies to specifying network scenarios described in section 1.4 and addressed throughout the thesis. The main reason for this are the simulation setups used in simulations of performances of MMP (and Mobile IP and Hierarchical Mobile IP) and Handover Management protocol, where the size of IP network and the number of attaching MHs correspond to typical layouts and populations encountered in campus-wide IP network environments. Additional performance analysis using the mathematical models in Chapter 4 gives broader performance characteristics than the ones limited by the simulation scenarios. The current state of the cellular telecommunications environments suggests significantly larger network sizes and population of terminals for which the conclusions based on the simulations may need to be re-examined. Hence, the mathematical models used in Chapter 4 could be generalised, adapted and extended for any mobility protocol and any network and population scenario. Additionally, a common simulation platform for validation of all

proposed protocols could be beneficial for completing the investigation. Although quite complex to create, a common simulation or test-bed platform for including protocols such as BCMP, MMP, Cellular IP, HAWAII, Hierarchical Mobile IP or any other relevant solution would further clarify the operational differences between the protocols. Alternatively, mathematical models could be used to provide such evaluation opportunity.

- **Potential possibilities for MMP in IPv6 should be further explored:** Although this document contains a description of a version of the protocol for IPv6 (see section 3.6), additional research is being carried out to enhance the protocol concerning the macro mobility section of the protocol, multiple Gateways, security, IP multicast routing, addressing and inclusion of relevant PDI Solutions such as the implementation of Handover Management PDI Solution.
- **TCP performances:** Performance of a connection oriented TCP transport protocol with IP mobility protocols could be examined, as this was not conducted in the thesis (some examples of TCP performances over MMP are given in [6][7]). The aim of such research could be: identifying effects of handover losses on TCP performance and proposing possible adjustments to TCP for supporting mobility scenarios.
- **Further application of the principles of Generic Mobility Design Model:** The model can be utilised for designing or refining other mobility protocols depending on some specific design principles and deployment scenarios not considered in the thesis. Hence, the model can facilitate development of new mobility solutions based on the appropriate selection of primary PDIs. The modularity principle of the model can be used to improve existing mobility protocols as exemplified in the case of MMP in Chapters 5 and 6. Such mobility solutions can be further evaluated using the general methods applied in Chapter 4 for validation and further performance evaluation of mobility protocols. A similar remark can also apply to Handover Management PDI developed in Chapter 6 for which similar evaluations can be conducted when integrated with other mobility protocols such as MMP.

REFERENCES

R1 Personal Publications Related to the Presented Work

[1] A. Mihailovic, M. Shabeer and A. H. Aghvami, "Multicast for Mobility Protocol (MMP) for emerging internet networks", The eleventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2000), Sep. 2000, London UK.

[2] A. Mihailovic, M. Shabeer and A.H. Aghvami, "Sparse mode multicast as a mobility solution for internet campus networks", The tenth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Conference Proceedings, Sep. 1999, Osaka Japan.

[3] A. Lopez, J. Manner, A. Mihailovic, H. Velayos, E. Hepworth and Y. Khouaja "Evaluation of Mobility and QoS interactions", Computer Networks, Volume 38, Issue 2, pages 137-163 The International Journal of Computer and Telecommunications Networking, 5 February 2002.

[4] A. Stephane, A. Mihailovic, A.H. Aghvami "Mechanisms and hierarchical topology for fast handover in wireless IP networks", IEEE Communications Soc Magazine, November 2000 pp112-115.

[5] P. Eardley, A. Mihailovic, T. Suihko, "A Framework for Evaluation of IP Mobility Protocols", The eleventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2000), September 2000, London UK (paper is submitted as a part of European Commission sponsored project Broadband Radio Access for IP-Based Networks - BRAIN, IST project IST-1999-10050).

[6] A. Delgado, A. Mihailovic, N. Georganopoulos, H. Aghvami, "Adaptation of Transport Protocols for an IP-Micromobility Scheme", International Conference on Telecommunication (ICC 2001), June 2001, Helsinki, Finnland.

[7] J. Corella, A. Mihailovic, N. Georganopoulos, A. H. Aghvami, "Analysis of Multicast for Mobility Protocol for IPv6 Networks", The Forth international Symposium on Wireless Personal Multimedia Communications (WPMC '01), September 20001, Aalborg, Denmark.

[8] A. Mihailovic, M. West, R. Hancock, P. Eardley, T. Suihko, "Experience of the BRAIN and MIND Projects in the Development of IP Mobility Solutions", Internet Draft (work in progress), draft-mihailovic-brain-mind-00.txt, September 2002.

R2 Other References

[9] J. Postel, "User Datagram Transport", RFC 768, August 1980.

[10] DARPA Internet Program, "Transmission Control Protocol, Protocol Specification", RFC 793, September 1981.

[11] J. Postel, "Internet Control Message Protocol", RFC 792, September 1981.

[12] S. Deering, "Host Extensions for IP Multicasting" RFC 1112, August 1989.

[13] W. Fenner, "Internet Group Management Protocol, Version 2", RFC 2236, November 1997.

[14] B. Cain, S. Deering, B. Fenner, I.Kouvelas, A. Thyagarajan, "Internet Group Management Protocol, Version 3", Internet Draft (work in progress), draft-ietf-idmr-igmp-v3-07.txt, March 2001.

[15] D. Waitzman, S. Deering and C. Partridge, "Distance Vector Multicast Routing Protocol", RFC 1075, November 1988.

[16] J. Moy, "Multicast routing extension for OSFP", RFC 1584, March 1994.

[17] S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C. Liu and L. Wei, "Protocol Independent Multicast (PIM): Protocol Specification", Internet Draft, January 1995.

[18] A. Ballardie, "Core Based Trees (CBT version 2) Multicast Routing ", RFC 2189, September 1997.

[19] A. Ballardie," Core Based Tree (CBT) Multicast Routing Architecture", RFC 2201, September 1997.

[20] C. Hedrick, "Routing Information Protocol", RFC 1058, June 1988.

[21] J. Moy, "Open Shortest Path Routing Version 2", RFC 1247, July 1991.

[22] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg, "SIP: Session Initiation Protocol", RFC 2543, March 1999.

[23] P. Mockapetris, "Domain Names – Concepts and Facilities", RFC 1034, November 1987.

[24] Droms R., "Dynamic Host Configuration Protocol", RFC 1541, October 1993.

[25] R. W. Stevens, "TCP/IP illustrated Volume 1 and 2", Addison Wesley.

[26] S.E. Deering, "ICMP Router Discovery Messages", RFC 1256, September 1991.

[27] D.B. Johnson and C.E. Perkins, "Route Optimisation in Mobile IP", Internet Draft (work in progress) draft-ietf-mobileip-optim-10.txt, November 2000.

[28] C. Perkins, ed., "IP Mobility Support", RFC 2002, October 1996.

[29] C. Perkins, "Mobile IP", IEEE Communication Magazine, May 1997 (contains Hierarchical Foreign Agents) or, C. Perkins, "Mobile-IP Local Registration with Hierarchical Foreign Agents", Internet Draft (work in progress), draft-perkins-mobileip-hierfa-00, February 1999.

[30] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regional Registration", Internet Draft (work in progress), draft-ietf-mobileip-reg-tunnel-04, March 2000.

[31] P. McCann. T. Hiller, J. Wang, A. Casati, C. Perkins, P. Calhoun, "Transparent Hierarchical Mobility Agents (THEMA)", Internet Draft (work in progress), draft-mccann-thema-00.txt, March 2000.

[32] K. El. Malki , N.A. Fikouras. S.R. Cvetkovic, "Fast Handoff Method for Real-Time Traffic over Scaleable Mobile IP Networks", Internet Draft (work in progress), draft-elmalki-mobileip-fast-handoffs-01.txt, Jun3 1999.

[33] C. Castelluccia, "A Hierarchical Mobile IPv6 Proposal", Technical Report No 0226 INRIA, November 1998.

[34] R. Caceres and V. Padmanabhan, "Fast and Scalable Handoffs for Wireless Internetworks", Proceedings of ACM Mobicom, November 1996.

[35] A. G. Valko, "Cellular IP - A New Approach to Internet Host Mobility," ACM Computer Communication Review, January 1999.

[36] R. Ramjee, T. La Porta, S. Thuel and K. Varadhan, “IP micro-mobility support using HAWAII”, Internet Draft, (work in progress), draft-ietf-mobileip-hawaii-00, June 1999.

[37] R. Ramjee, T. La Porta, and L. Li, "Paging support for IP mobility using HAWAII", Internet Draft (work in progress), draft-ietf-mobileip-paging-hawaii-00.txt, June 1999.

[38] S. Seshan, H. Balakrishnan and R. H. Katz, “Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience”, ACM/Baltzer Journal on Wireless Networks, 1995.

[39] J. Mysore and V. Bharghavan, “A New Multicasting-based Architecture for Internet Host Mobility”, Proceeding of ACM Mobicom, September 1997.

[40] C. Tan, S. Pink, and K. Lye, "A Fast Handoff Scheme for Wireless Networks", In Proceedings of the Second ACM International Workshop on Wireless Mobile Multimedia, ACM, August 1999.

[41] A. O'Neill, G. Tsirtsis, and S. Corson, "Edge Mobility Architecture", Internet Draft (work in progress), draft-oneill-ema-01.txt, March 2000.

[42] V. Park and S. Corson, “Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification”, Internet Draft (work in progress) , draft-ietf-manet-tora-spec-02.txt, October 1999.

[43] C. Perkins, “IP Encapsulation within IP”, RFC 2003, May 1996.

[44] C. Perkins, “Minimal Encapsulation within IP”, RFC 2004, May 1996.

[45] S. Hanks, “Generic Routing Encapsulation (GRE)”, RFC 1701, October 1994.

[46] Y. Rekhter, T. Li, “A Border Gateway Protocol 4 (BGP-4)”, RFC 1771, March 1995.

[47] The BRAIN project Public Deliverable 2.2, "BRAIN architecture specifications and models, BRAIN functionality and protocol specification", Broadband Radio Access IP-based Networks – BRAIN, IST-1999-10050 BRAIN, http://jungla.dit.upm.es/~ist-brain/deliverables/BRAIN%20Del%202.2.pdf, March 2001.

[48] R. Hancock, H. Aghvami, M. Kojo, M. Liljeberg, "The Architecture of the BRAIN network Layer", Proceedings of IST Mobile Communications Summit 2000, October 2000, pp 581-586.

[49] A. O'Neill, G. Tsirtsis, and S. Corson, "Generalized IP Handoff", Internet draft draft-oneill-craps-handoff-00.txt, August 2000.

[50] A. O'Neill, G. Tsirtsis, and S. Corson, "EMA Enhanced Mobile IPv6/IPv4", Internet draft, draft-oneill-ema-mip-00.txt, July 2000.

[51] K. El-Malki and H. Soliman, "Fast Handoffs in Mobile IPv4", Internet draft, draft-elmalki-mobileip-fast-handoffs-03.txt, September 2000.

[52] K. El-Malki and H. Soliman, "Fast Handoffs in MIPv6", Internet draft, draft-elmalki-handoffsv6-01.txt, November 2000.

[53] R. Koodli and C.Perkins, "A Framework for Smooth Handovers with Mobile IPv6", Internet draft, draft-koodli-mobileip-smoothv6-01.txt, November 2000.

[54] P. Calhoun, et. al., "Foreign Agent Assisted Hand-off", Internet draft, draft-calhoun-mobileip-proactive-fa-03.txt, November 2000.

[55] R. Koodli and C. Perkins, "Fast Handovers in Mobile IPv6", Internet draft, draft-koodli-mobileip-fastv6-01.txt, October 2000.

[56] Karim El Malki (editor), "Low latency Handoffs in Mobile IPv4", Internet draft, draft-ietf-mobileip-lowlatency-handoffs-v4-00.txt, February 2001.

[57] C. Perkins, "Fast Handovers for Mobile IPv6", Internet draft , draft-ietf-mobileip-handover-00.txt, November 2000.

[58] G. Tsirtsis, "Fast Handovers for Mobile IPv6", Internet draft , draft-designteam-fast-mipv6-01.txt, February 2001.

[59] G. Tsirtsis, A. Yegin, C. Perkins, G. Dommety, K. El-Malki, M. Khalil , "Fast Handover for Mobile IPv6", Internet draft (work in progress) , draft-ietf-mobileip-fast-mipv6-03.txt, November 2001.

[60] K. El-Malki, P. Calhoun, t. Hiller, J. Kempf, P. McCann, A. Singh, H. Soliman, S. Thalanany, "Low Latency Handoff in Mobile IPv4", Internet Draft (work in progress), draft-ietf-mobileip-lowlatency-handoffs-v4-03.txt, May 2001.

[61] C. Keszei, N. Georganopoulos, Z. Turanyi, A. Valko, "Evaluation of the BRAIN Candidate Mobility Management Protocol", IST Summit Barcelona, September 2001.

[62] ETSI BRAN and HIPERLAN, HIPERLAN Type 2, www.etsi.org/bran.

[63] Sarikaya B. Haverinen H., Malinen J.T., Magret V., "Mobile IPv6 Regional Paging", Internet Draft (work in progress), draft-sarikaya-mobileip-hmipv6rp-00.txt, November 2000.

[64] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource RESerVation Protocol (RSVP) – Version 1 Functional Specification", RFC 2205, September 1997.

[65] M. Mouly, M.-B. Pautet, "The GSM System for Mobile Communications", Cell & Sys, 1992.

[66] WAP Forum, "Wireless Application Architecture Specification", URL: http://www.wapforum.org , April 1999

[67] H. Granbohm, J. Wiklund, "GPRS General Packet Radio Service", Ericsson Review, No.2, 1999.

[68] S. Da Silva, B. Arroyo-Fernandez, B. Barani, J. Pereira, D. Ikonomou, "Evolution Towards UMTS", European Commission – DG XIII-B.4, 1996.

[69] URL: http://www.3gipgroup.org

[70] M. Annoni, R. Hancock, T. Paila, E. Scarrone, R. Tonjes, L. Dell Uomo, D. Wisely, R. Mort, "Radio Access Networks beyond the 3rd Generations: A First Comparison of Architectures from 4 IST Projects" IST Summit, September 2001.

[71] C. Perkins, P. Calhoun, "AAA Registration Keys for Mobile IP", Internet Draft (work in progress) draft-ietf-mobileip-aaa-key-08.txt, July 2000.

[72] J. Kempf, "Problem Description: Reasons For Performing Context Transfers Between Nodes in an IP Access Network Context and Micro-mobility Routing Working Group", Internet Draft (work in progress) draft-ietf-seamoby-context-transfer-problem-stat-03.txt, April 2002.

[73] S.Deering, R. Hinden, "Internet Protocol Version 6 (IPv6) Specification", RFC 2460, December 1998.

[74] D. Johnson, C. Perkins, "Mobility Support in IPv6", Internet Draft (work in progress) draft-ietf-mobileip-ipv6-15.txt, July 2001.

[75] S. Thomson, T. Narten, "IP Stateless Address Autoconfiguration", RFC 2462, December 1998.

[76] T. Narten, E. Nordmark, W. Simpson, "Neighbour Discovery Protocol for IP Version 6 (IPv6)", RFC 2461, December 1998,

[77] J. Bound, C. Perkins, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", Internet Draft (work in progress) draft-ietf-dhc-dhcpv6-21.txt, November 2001.

[78] R. Hinden. S. Deering, "IP Version 6 Addressing Architecture", RFC 2373, July 1998.

[79] Z. Shelby, D. Gatzounas, A. Campbell, C. Wan, "Cellular IP", Internet Draft (work in progress) draft-shelby-seamoby-cellularipv6-01.txt, July 2000.

[80] A. Conta, S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6)", RFC 2463, December 1998.

[81] C. Barakat, E. Altman, and W. Dabbous, "On TCP Performance in a Heterogeneous Network : A Survey", IEEE Communication Magazine, vol. 38, no. 1, pp. 40-46, January 2000.

[82] R. Ramjee, K. Varadhan, L. Salgrelli, S.R. Thuel, S. Wang, T. Porta, "HAWAII: A Domain-Based Approach for Supporting Mobility in Wide Area Wireless Networks", IEEE/ACM Transactions on Networking, Vol. 10, No. 3, June 2002.

[83] A. T. Campbell, J. Gomez, S. Kim, C. Wan, Z. Turanyi, A. Valko, "Comparison of IP Micromobility Protocols", IEEE Wireless Communications Magazine, February 2002.

[84] A.T. Campbell, Gomez, J., Kim, S., Turanyi, Z., Wan, C-Y. and A, Valko "Design, Implementation and Evaluation of Cellular IP", IEEE Personal Communications Magazine, Special Issue on IP-based Mobile Telecommunications Networks, June/July 2000.

[85] P. De Silva, H. Sirisena, "A Mobility Management Protocol for IP-based Cellular Networks", IEEE Wireless Communications Magazine, June 2002.

[86] A. G. Valko, J. Gomez, S. Kim, A. T. Campbell, "On the Analysis of Cellular IP Access Networks", IFIP Sixth International Workshop on Protocols for High Speed Networks (PfHSN'99), Salem Massachusetts, August 1999.

[87] DARPA Internet Program, "Internet Program, Protocol Specification", RCF 791, September 1981.

[88] K. Fall, S. Floyd, "Simulation-based Comparisons of Tahoe, Reno and SACK TCP", Computer Communication Review, Vol.26, No.3, July 1996.

[89] The MIND project Public Deliverable 2.2 (Core and Annex parts), "MIND protocols and mechanisms specification, simulation and validation", Mobile IP-based Network Developments – MIND, IST-2000-28584 MIND http://www.ist-mind.org/deliverables.htm, November 2002.

R3 Other personal publications related to the presented work

[90] J. Manner, A. López, Andrej Mihailovic, H. Velayos, E. Hepworth, Y. Khouaja, "Evaluation of Mobility and QoS Interactions", 2nd International Workshop on Broadband Radio Access for IP-based Networks, Yokosuka Research Park, Yokosuka, Japan, February 2001.

[91] A. López, J. Manner, A. Mihailovic, H. Velayos, E. Hepworth and Y. Khouaja, " Study on QoS Provision for IP-based Radio Access Networks", 2001 Tyrrhenian International Workshop on Digital Communications Taormina, Italy, September 2001. (Invited paper).

[92] K. David, P. Eardley, D. Hetzer, A. Mihailovic, T. Suihko, M. Wagner, "A first evaluation of IP based network architectures", 2nd International Workshop on Broadband Radio Access for IP-based Networks, Yokosuka Research Park, Yokosuka, Japan, February 2001.

[93] A. Delgado, A. Mihailovic, N. Geoganopoulos, A. H. Aghvami, "Evaluating the performance of Transport Protocols over MMP", Proceedings of the London Communications Symposium (LCS2000), London, UK, September 2000.

[94] A. Mihailovic, G. Leijonhufvud, T. Suihko, "Providing Multi-homing Support in IP Access Networks", The thirteenth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002), Lisbon, Portugal, September 2000.

[95] A. Mihailovic, T. Suihko, M. West, "Aspect of Multi-homing in IP Access Networks", IST Mobile and Wireless Summit , Thessalonici, Greece, 2002.

[96] V. Friderikos, A. Mihailovic, K. Samandis, A.H. Aghvami, "Analysis of Cross Issues Between QoS Routing and μ-Mobility Protocols", Proceeding of Tenth Meeting of World Wide Research Forum (WWRF10), New York, USA, 2003.

[97] V. Friderikos, A. Mihailovic, A.H. Aghvami, "Analysis of Cross Issues Between QoS Routing and μ-Mobility Protocols", to appear in IEE Proceedings Communications, Special Issues on Internet Protocols, Technology and Applications (VoIP), Vol. 151, No. 3, 2004.

APPENDIX ONE

A1 BRAIN Handover Protocol Specifications

A1.1 Introduction

The following sections present the exact sequence of the message transfer in the unplanned and planned handover cases explained in sections 6.2.1 and 6.2.2. This particular specification is taken from BRAIN results shown in [47] (see Appendix 2 from further explanation of the results of the project and author's involvement). The details of the message syntax are not presented as well as the dependency on the particular version of IP and the signalling formats of messages: ICMPv4, ICMPv6 [80], IPv6 Destination Options and Mobile IP signalling. The handover protocol aims to achieve compatibility between both versions of IP. Thus, equivalence between IPv4 and IPv6 is assumed: Agent versus Router Advertisements and Binding Update and Registration Request are regarded as functionally equal. The most significant constraint for an IPv4/IPv6 generic handover protocol is the address acquisition/autoconfiguration differences between the versions. However, this is not

an issue in this handover protocol since, as mentioned in Chapter 6, it is intended for static/semi-static care-of-address in a *Localised Enhanced-Routing Scheme* scenario where there is no change of address for every handover.

A1.2 Planned Handover

The planned handover presented in Chapter 6 should have the following steps (see Figure A1.1):

1. A link-trigger occurs at the MH or at the old BS. This initiates the handover (a particular adaptation of this handover algorithm with the underlying link layer, HIPERLAN triggers, can be found in [47]).
2. If the trigger occurs at the MH it may solicit (*CAR Solicitation*) information about candidates new BSs from the old BS (*CAR Advertisements*). This is an optional step and may be avoided if the MH already identifies one or more candidates. Then a *Host Handover Request* is sent to the old BS directly.
3. The old BS responds to the *CAR Solicitation* with a *CAR Advertisement*, which contains identification of the candidates and may additionally contain link layer parameters for accessing them (the old BS is assumed to have awareness of the level of "handover willingness" of the candidates).
4. Regardless of whether the MH has decided on the candidate(s) based on its own mechanism (radio signal measurement or policy decisions) and/or it has consulted the *CAR Advertisement* from the old BS, it needs to specify the list of (or one) candidates in the *Host Handover Request*.
5. The old BS relays the request to the candidates specified by the MH. The *Host Handover Request* sent by the MH is used to construct the *BS Handover Request*, which is sent from the old BS to the chosen candidates. The *BS Handover Request* contains the MH's general context: identification, IP address, link-layer address, session keys and other context (can also be used as a generic container

for other contexts: QoS, header compression...). Alternatively, MH could send the message to the new BS straight away this depending on the implementation preferences.

6. Candidates reply to the *BS Handover Request* with a *BS Handover Reply* to indicate whether they are accepting or declining the handover.
7. The old BS informs the MH about the success of the handover request by sending a *Host Handover Reply* to the MH. At this point, the old BS starts tunnelling packets to the new BS (or candidates if MH decides to have more than one candidate at this stage).
8. The MH attaches to the new BS. The exchange of Router/Agent[1] Solicitation and Advertisement messages indicates a completion of registrations (regarding the link layer this network layer step may be preceded by associated link layer procedures for obtaining transmission resources: medium contention procedures in random access links or channel reservation in connection oriented link layers). The new BS can then start forwarding packets received from the old BS. The new BS should buffer the packets until the MH performs the registration. If the MH establishes a link through the link layer procedure, which unambiguously indicates that the MH has connected to the new BS, the new BS can use this to start forwarding. Otherwise it may defer forwarding until registration.
9. The MH registers by sending a *Registration Request*[2].
10. The new BS generates a *Path Update Request*[3] according to the specification of the rest of the IP mobility protocol (or it may relay the Registration Request) towards a "cross-over" router.
11. The new BS receives an indication that Path Updates have been performed successfully by typically receiving an acknowledgment in the form of the *Path*

---

[1] IPv6/IPv4.
[2] Adopted from IPv4 and corresponds to Binding Update in Mobile IPv6.
[3] This a general term representing any Path Update message of a mobility setup (CBT Join Request in MMP).

Update Reply from a router in the network (in MMP this is the CBT Join Ack, in BCMP the login reply from the ANP).

12. Finally, the new BS relays the *Registration Reply* (or any other equivalent message) to the MH.

After the downstream routing path has been diverted to the new BS, packet forwarding at the old BS can be terminated. This can be achieved by a timeout mechanism at the old BS or by an explicit signal from a "cross-over" router or the new BS depending on the mobility protocol (see sections 6.3.1 and 6.3.2 for the ways in which MMP and BCMP solve this particular issue). This signal is not shown in Figure A1.1.

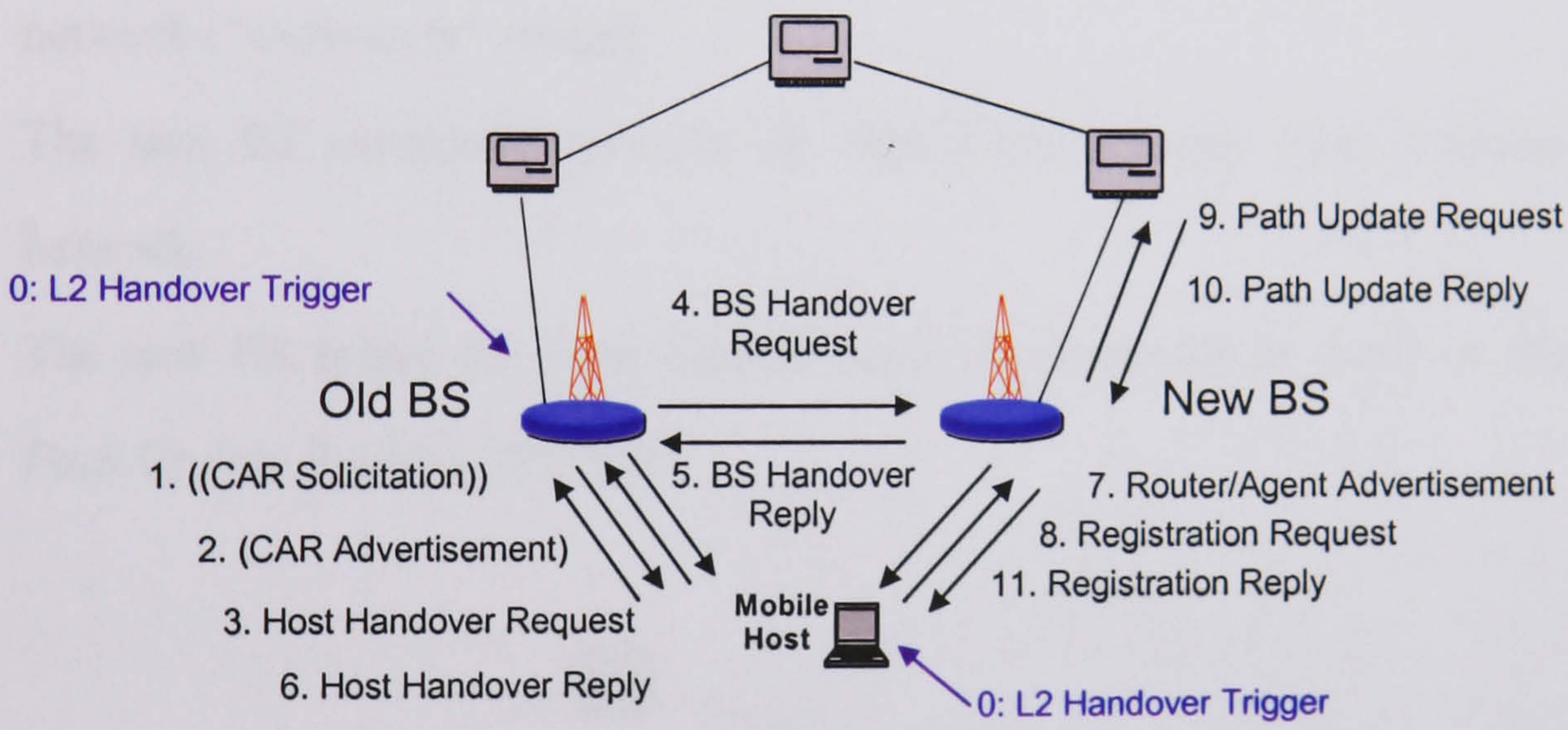


Figure A1.1. Planned handover

A1.3 Unplanned Handover

The unplanned handover described in Chapter 6 should have the following steps (see Figure A1.2):

1. A link-layer trigger (either an indication that a handover is needed or of a loss of connection to the old BS, i.e. failure of the planned handover)
2. The MH establishes a link with the new BS.

3. Exchange of *Router/Agent Solicitation* and *Advertisements*. There may only be an *Advertisement* sent by the new BS.
4. The MH sends a *Registration Request* (*Binding Update*) to the new BS. The request indicates the identification of the old BS along with any possible context that may be transferred.
5. The new BS sends a *BS Handover Request* to the old BS and may request the transfer of any possible context to be transferred.
6. The old BS responds with a *BS Handover Reply* along with any required context and starts tunnelling the packets to the new BS.
7. The new BS performs the Path Updates by generating a Path Update Request (or relays the Registration Request/Binding Update) towards the inside of the network ("cross-over" router).
8. The new BS eventually receives the Path Update Reply from a router in the network.
9. The new BS relays the *Path Update Reply* (as *Registration Reply* or the actual *Path Update Reply)* to the MH.

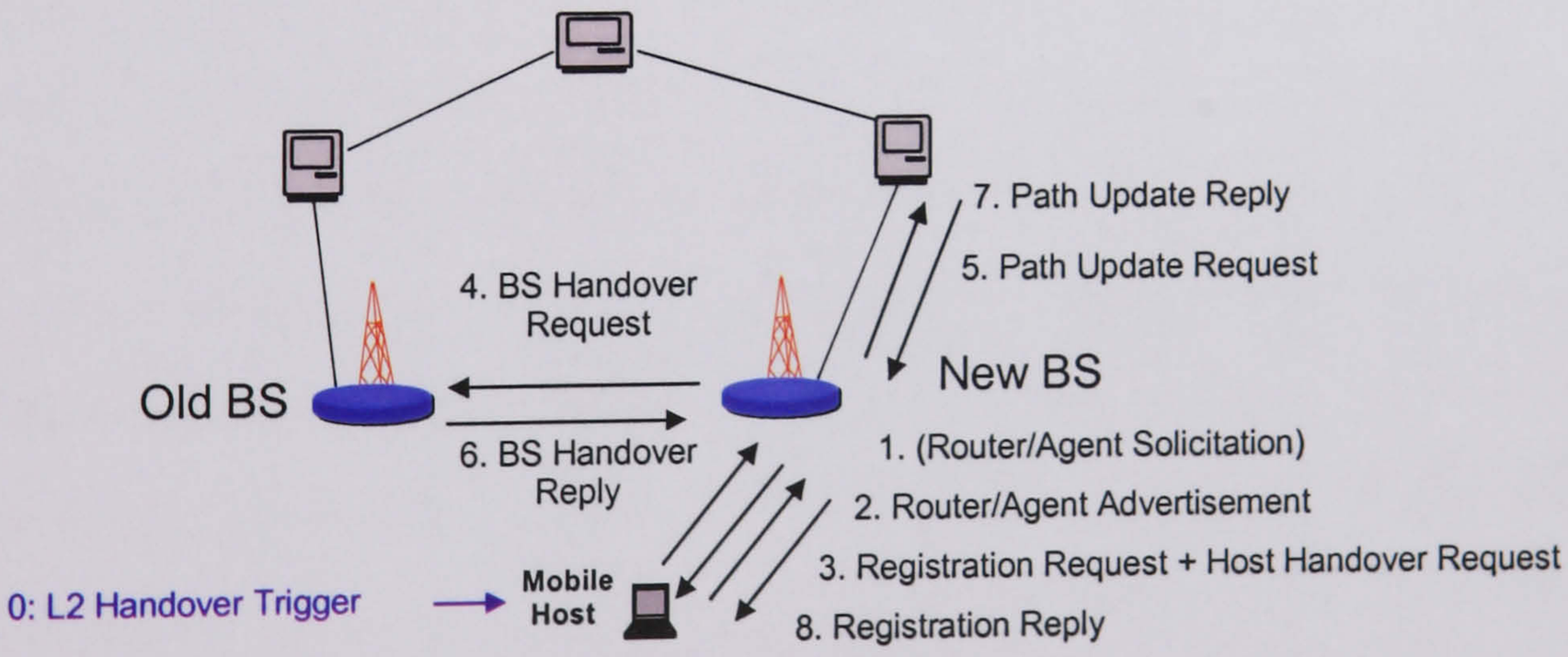


Figure A1.2. Unplanned handover

(Note: In specific security scenarios the new BS may perform the path updating part of the mobility protocol and consultation with the old BS simultaneously if the MH's Registration Request/Binding Update can be authenticated without first retrieving the MH's session key from the old BS)

APPENDIX TWO

A2 BRAIN/MIND Project Details

A2.1 Summary of the BRAIN Project

The Broadband Radio Access for IP-based Networks – BRAIN [47] (IST-1999-10050 BRAIN) project dealt with development of IP-based architecture for IP access networks (project consortium: Siemens AG, British Telecommunications PLC, Agora Systems S.A., Ericsson Systems AB, France Télécom S.A., INRIA, King's College London, Nokia Corporation, NTT DoCoMo Inc, Sony International (Europe) GmbH, T-Systems Nova GmbH, University of Madrid and Infineon Technologies AG). Extensive research into IP mobility protocols was one of the main activities of the project. The BRAIN Access Network (AN) is defined as static IP access network with wired internal infrastructure and heterogeneous wireless access for MHs. IP functionality is contained in the entire system from gateways to Access Routers (ARs) where MHs are assumed to run IP protocols with additional capabilities required to take advantage of the connectivity offered by the AN.

The basic goals of the BRAIN project in the area of the access network have been to design an IP-based AN which supports new air interfaces (e.g. Wireless LANs), adding functionality to allow them to complement current access systems. The main functions of this AN include IP-based micro-mobility handling, and also quality of service handling to provide seamless service provision and QoS adaptation in the face of, for example, radio signal deterioration or lower bandwidth on hand-over. The conceptual relationship of the BRAIN AN to other networks is shown in Figure A2.1.

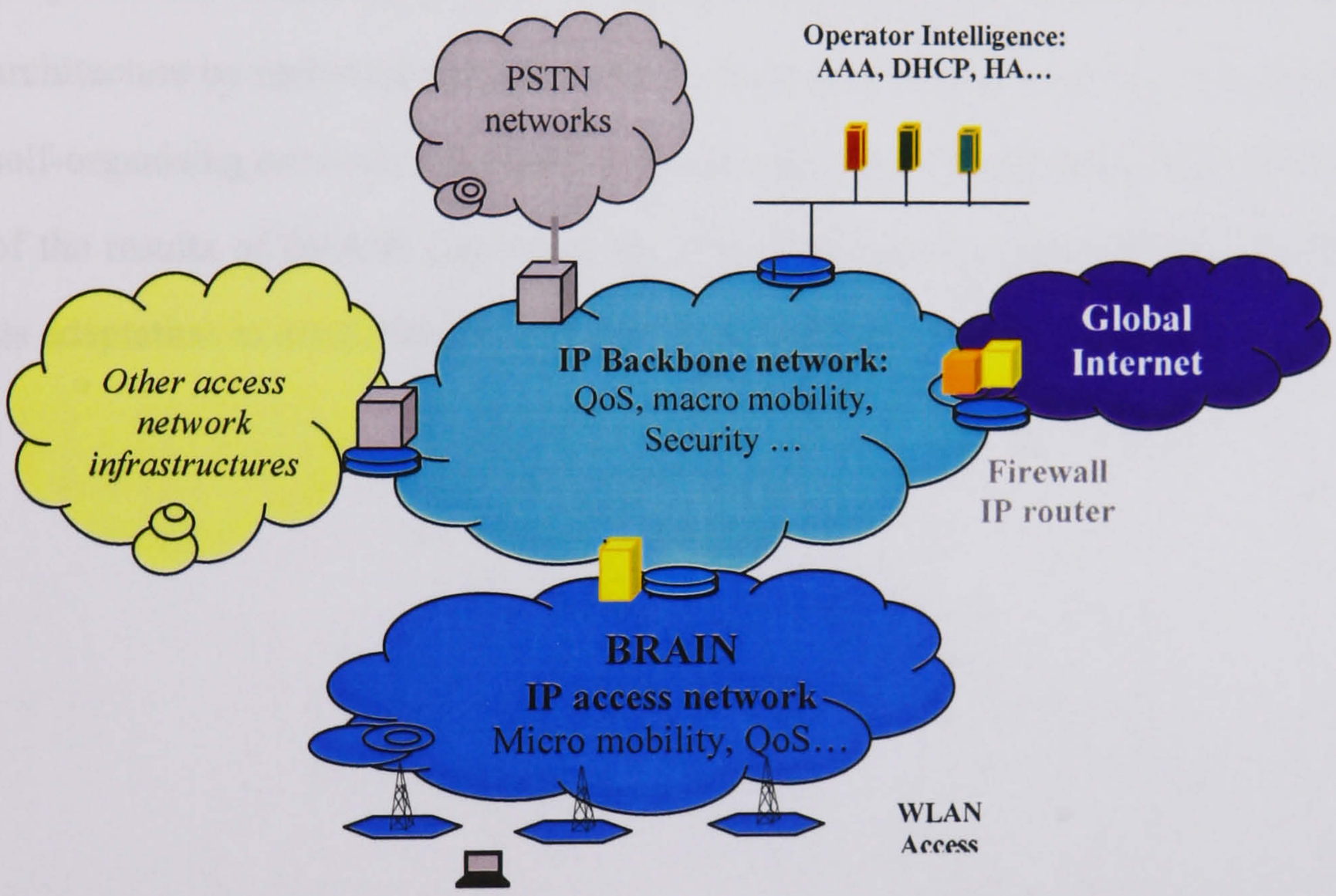


Figure A2.1. BRAIN network and its relation with other networks

In terms of layering, the BRAIN access network is restricted to pure transport of IP packets with assured QoS, security, and seamless handover (terminal mobility); all higher layer functions are transparent. While first aimed at support for Hiperlan/2 and other WLAN systems, the BRAIN layer is supposed to be adaptable to the case of IP over other air interfaces. This is achieved by using an enhanced IP layer in terminals and in some parts of the access network infrastructure, which uses the services

provided at a generic 'IP to Wireless' (IP_2W) service interface. Adaptations or enhancements to support the IP_2W are provided in an air interface-specific convergence layer below it. Much more detailed information about the project result is given in Deliverable 2.2 [47] which deals with network layer issues.

A2.2 MIND project – BRAIN follow-up

The Mobile IP-based Network Development – MIND project (MIND IST2000-28584) [89] was the follow-up project to BRAIN continuing the development of the AN architecture by including some new network scenarios of ad-hoc networks, mobile and self-organising networks. The MIND project also dealt with improvements and testing of the results of BRAIN (including the IP mobility protocol developed - BCMP) and its adaptation to multi-homing and MIND network scenarios.

APPENDIX THREE

A3 Depiction of Protocol Steps for Multicast for Mobility Protocol

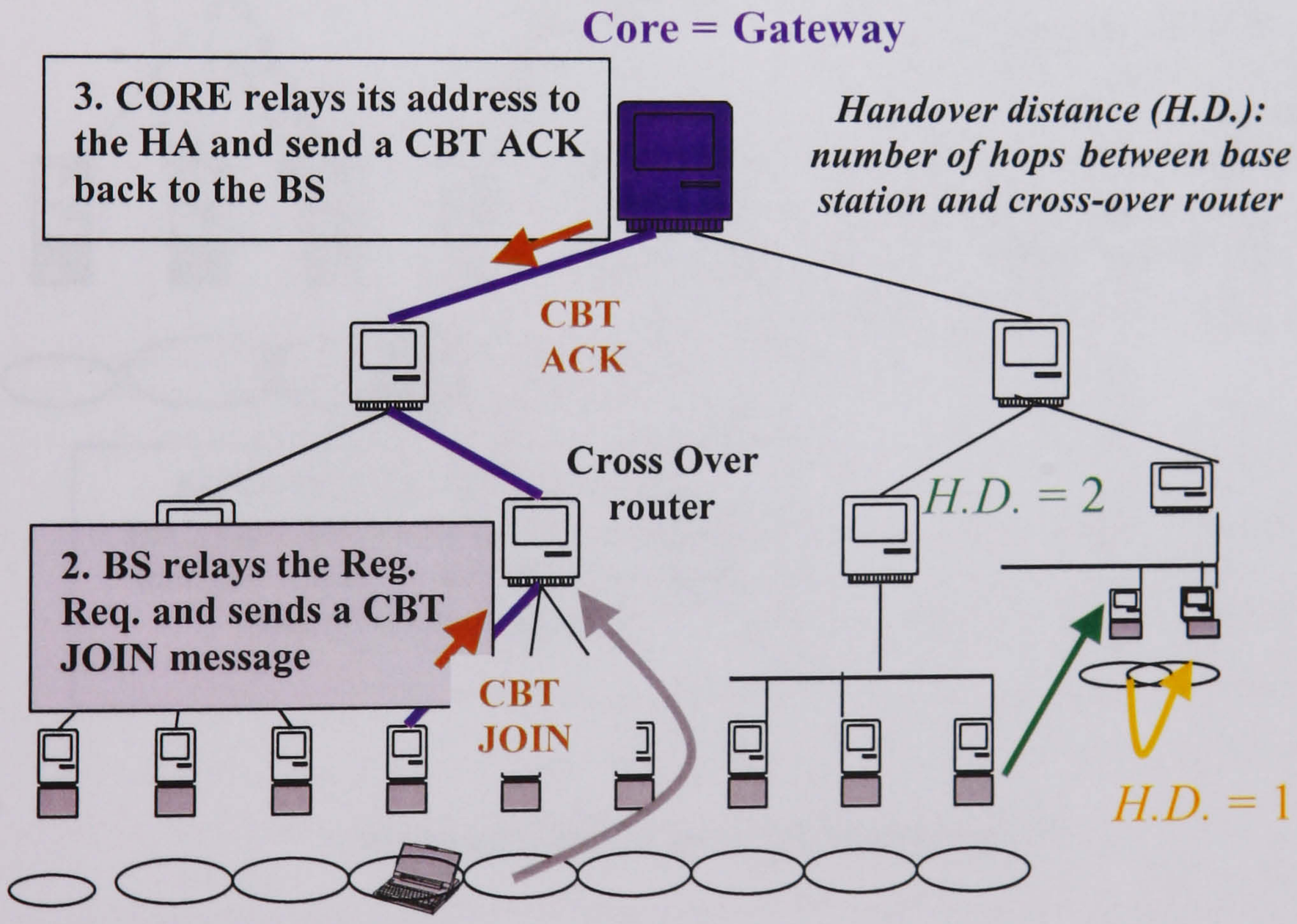


Figure A3.1. Handovers in MMP

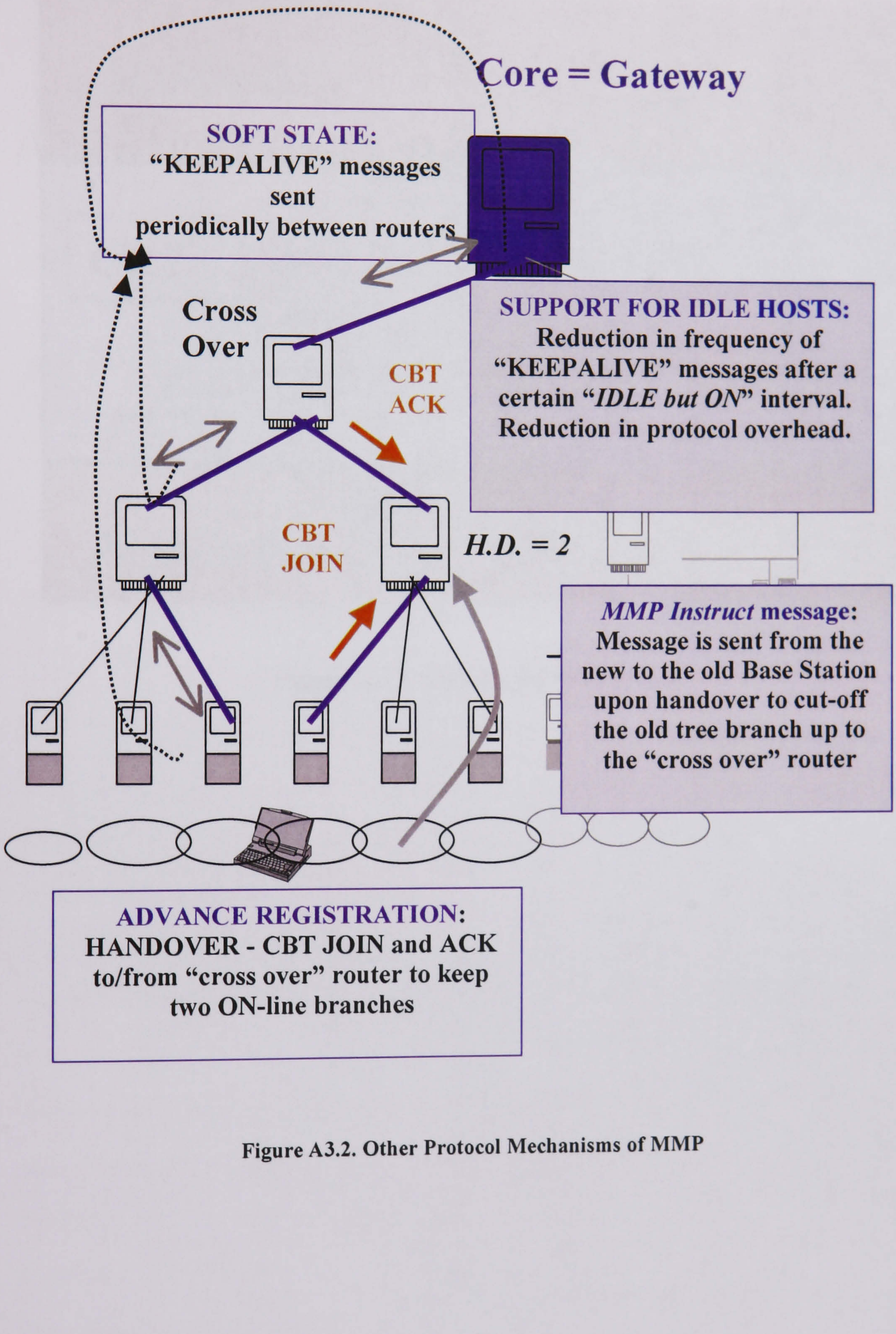


Figure A3.2. Other Protocol Mechanisms of MMP

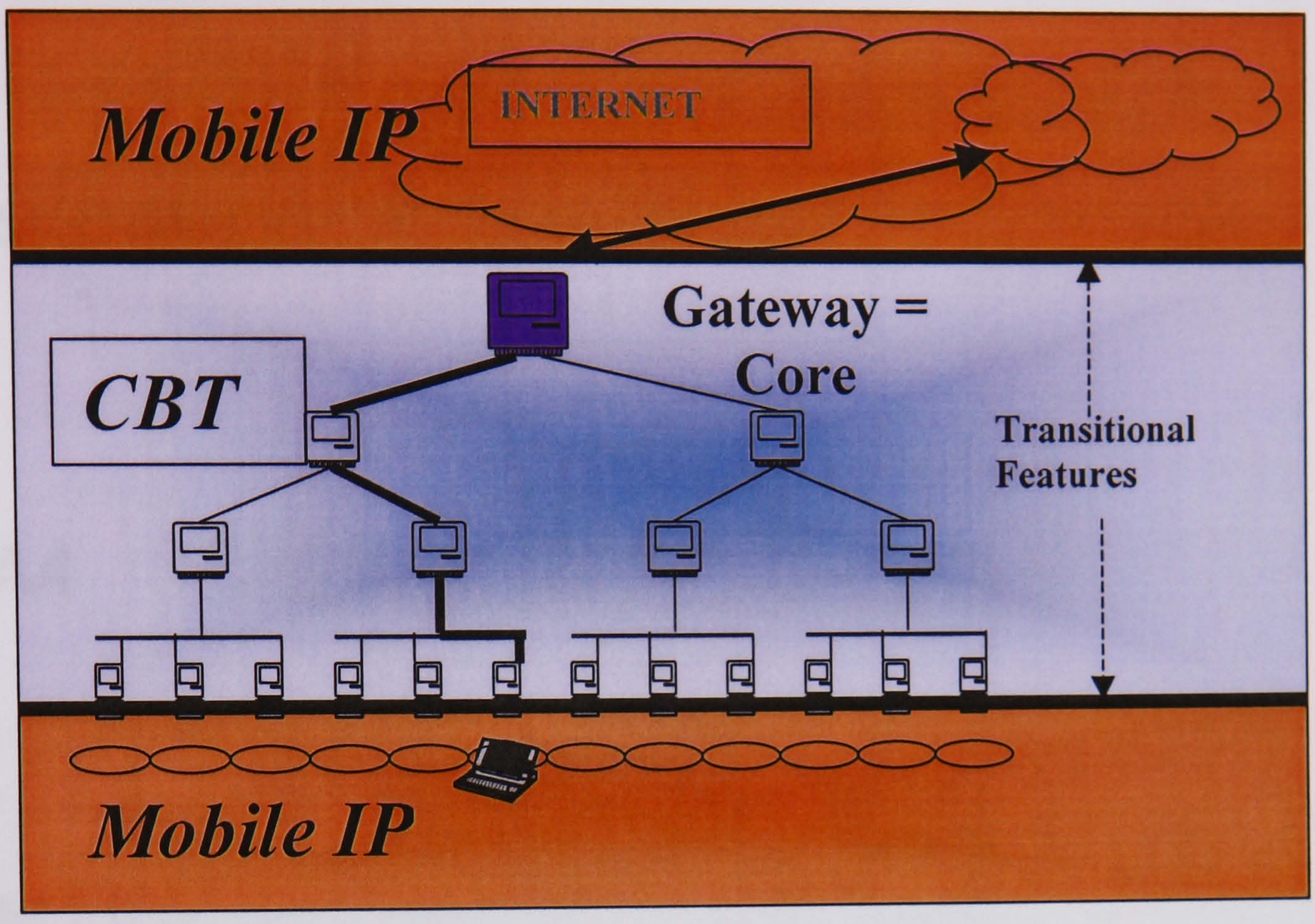


Figure A3.3. General Setup of MMP

APPENDIX FOUR

A4 Further Ns-2 Simulations

A4.1 Description of Network Simulator 2 (ns-2)

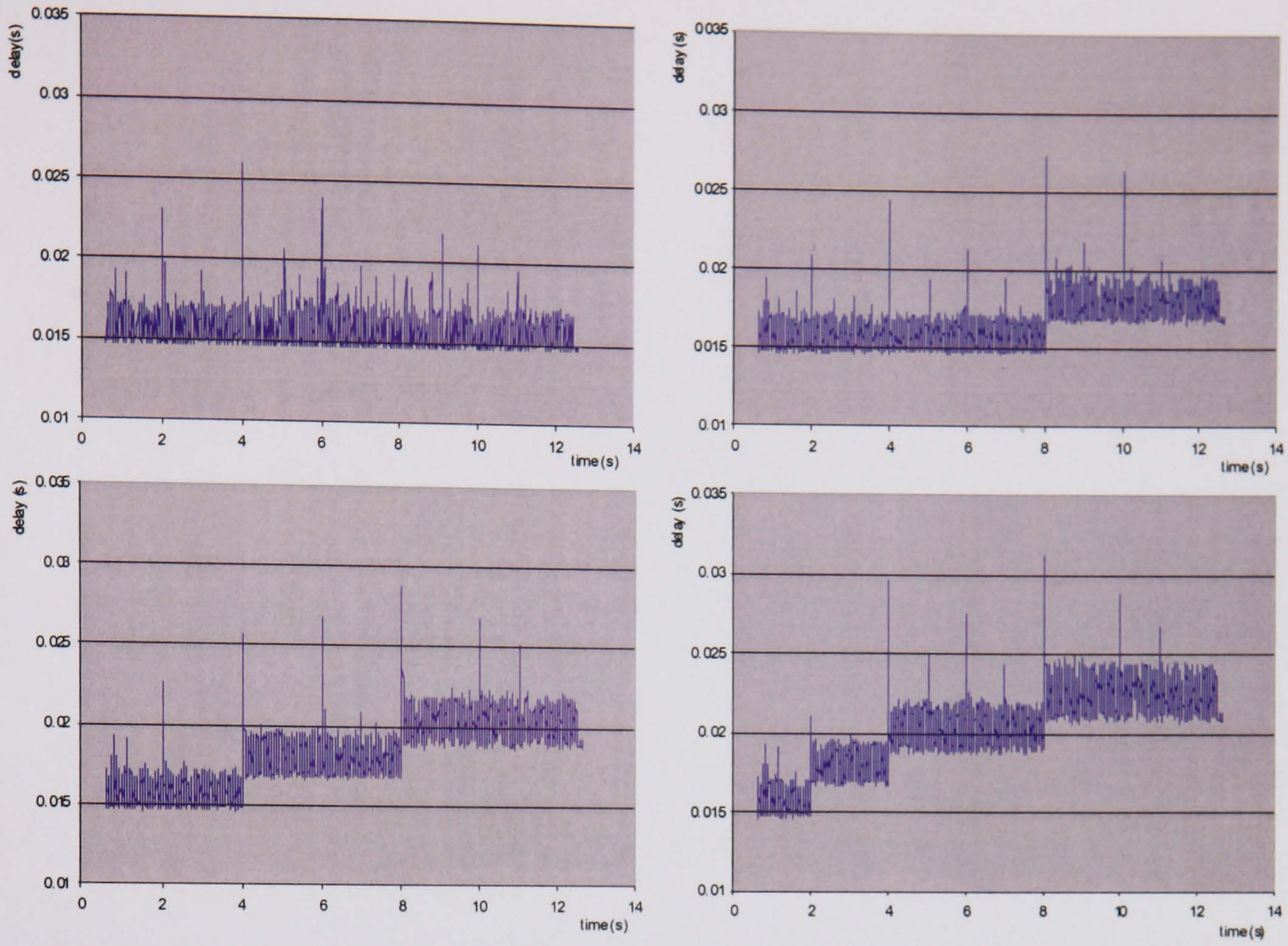
This chapter contains additional supporting simulation results related to the text in the thesis and simulations results presented in Chapter 4 and Chapter 6. The simulation tool used for the results is Network Simulator version 2 – ns-2. Ns-2 is a free network simulation program that is obtainable from the World Wide Web and is compatible with a number of operating systems. The tool has functionality for simulating different network topologies and traffic models. Ns-2 also has an open architecture that allows users to add new functionality. Ns-2 has been developed at the Lawrence Berkeley National Laboratory (LBNL) of the University of California, Berkeley (UCB). The extensibility of ns makes the tool very dynamic with frequent changes and widely available libraries of protocols. Ns-2 can be classified as an event-driven network simulator. It is build upon an extensible background engine implemented in C++ that uses OTcl (an object oriented version of Tool Command Language - Tcl) as the command and configuration interface. Thus, the entire software hierarchy is written

in C++, with OTcl used as a front end. The core of the majority of simulators, including ns-2, is a discrete event processor. The accuracy, performance and scaling are important to these simulators, hence several complementary steps are taken to improve upon them. One of these steps was to extend the event processor with analytic models of traffic flow or queuing behaviour for better performance or accuracy. Other ways of improving the performance and accuracy are parallel and distributed simulations. Network simulator (ns), network animator (nam) visualisation tool and topology generators provides several critical innovations that broaden the range of conditions under which existing and proposal of new protocols can be evaluated while making this experimentation tractable. Furthermore, several engineering issues have substantial impact on a simulator's usability. One of these issues is the availability of a wide range of protocol modules in the simulator.

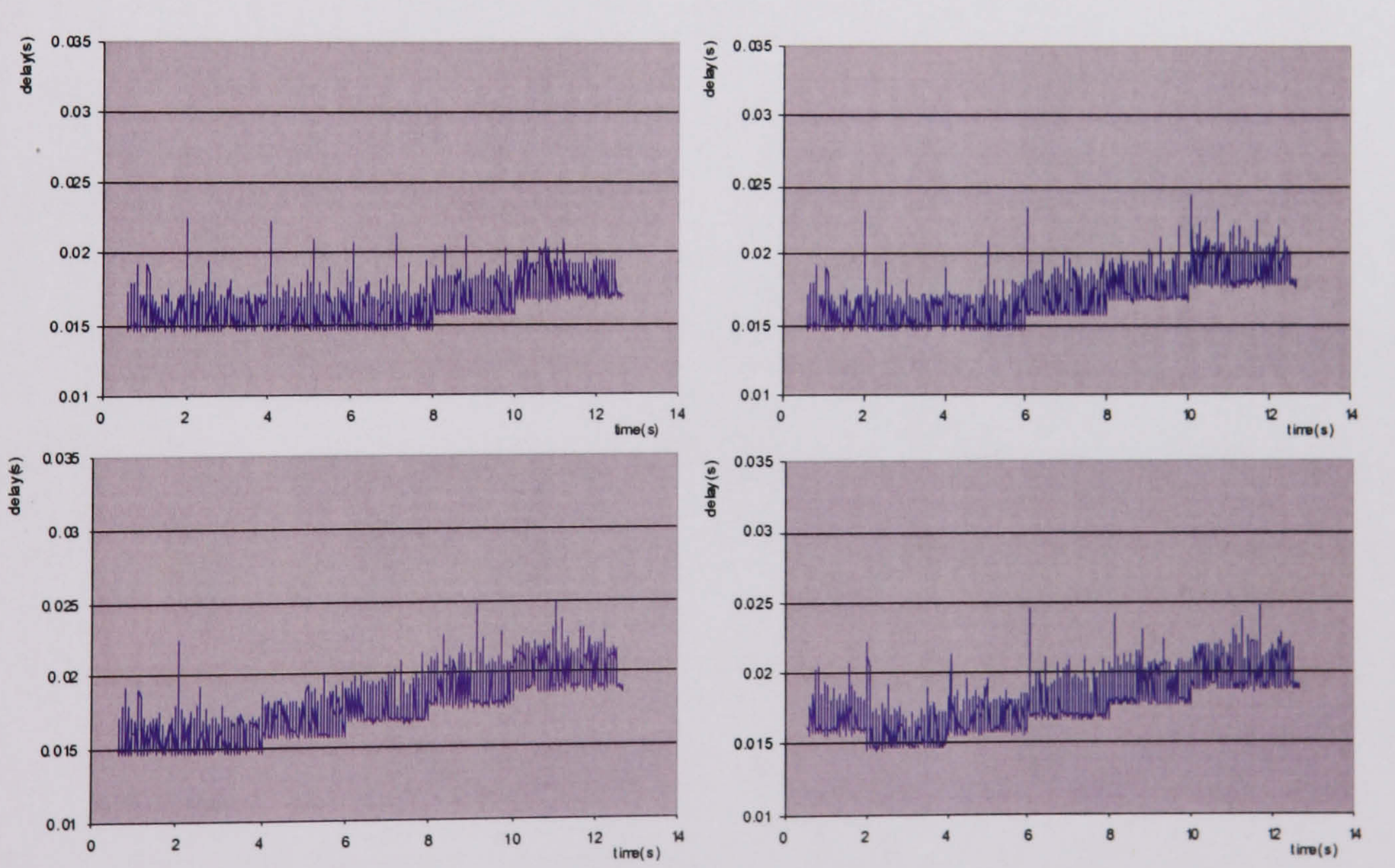
This public availability of ns-2 modules was the main reason why simulations presented in this chapter and Chapter 6 are performed by the ns-2 and not OPNET models as in Chapter 3. Since all simulations results apart from the ones presented in Chapter 4 are results of joint project efforts in BRAIN/MIND projects (see section 1.6 for author's involvement in the projects), the publicly available ns-2 was a more appropriate choice rather than the professional and licence-requiring OPNET Modeller, which the author of the thesis separately used for results presented in Chapter 4.

A4.2 BRAIN/MIND Handover Management (BCMP) Simulations in ns-2: Unplanned Handover

The following graphs support the description of the Handover Management protocol and its performance presented in Chapter 6 and are taken from BRAIN/MIND simulation results shown in [89] for the unplanned handover.



Graph A4.1: Unplanned Handover: Packet End-to-end Delays – Hierarchical Topology Configurations A, B, C, D (source [89])



Graph A4.2: Unplanned Handover Packet End-to-end Delays – Partial Mesh Topology Configurations A, B, C, D (source [89])